DOCUMENT RESUME

ED 390 082 CS 509 096

AUTHOR	Fowler, Carol A., Ed.
TITLE	Speech Research Status Report, January-June 1994.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, MD.
REPORT NO	SR-117-118
PUB DATE	94
CONTRACT	DBS-9112198; HD-01994; MH51230
NOTE	239p.; For the July-December 1993 report, see ED 378 624.
PUB TYPE	Collected Works - General (020) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC10 Plus Postage.
DESCRIPTORS	Adults; Communication Research; Elementary Secondary Education; Higher Education; Infants; *Language Acquisition; Language Research; Literacy; *Music; North American English; Reading; *Speech Communication; *Vowels
IDENTIFIERS	Pianos; *Speech Research

ABSTRACT

This publication (one of a series) contains 14 articles which report the status and progress of studies on the nature of speech, instruments for its investigation, and practical applications. Articles include: "The Universality of Intrinsic F0 of Vowels: (D. H. Whalen and Andrea G. Levitt); "Intrinsic F0 of Vowels in the Babbling of 6-, 9-, and 12-Month-Old French- and English-Learning Infants" (D. H. Whalen and others); "Knowledge from Speech Production Used in Speech Technology: Articulatory Synthesis" (Richard S. McGowan); "Nonsegmental Influences on Velum Movement Patterns: Syllables, Sentences, Stress, and Speaking Rate" (Rena A. Krakow); "Articulatory Organization of Mandibular, Labial, and Velar Movements during Speech" (H. Betty Kollia and others); "An Acoustic and Electropalatographic Study of Lexical and Post-Lexical Palatalization in American English" (Elizabeth C. Zsiga); "The Discriminability of Nearly Merged Sounds" (Alice Faber and Marianna Di Paolo); "The Role of Fundamental Frequency in Signaling Linguistic Stress and Affect: Evidence for a Dissociation" (Gerald W. McRoberts and others); "Orthographic Representation and Phonemic Segmentation in Skilled Readers: A Cross-Language Comparison" (Ilana Ben-Dror and others); "Expressive Timing in Schumann's 'Traumerei': An Analysis of Performances by Graduate Student Pianists" (Bruno H. Repp); "Quantitative Effects of Global Tempo on Expressive Timing in Music Performance: Some Perceptual Evidence" (Bruno H. Repp); "Detectability of Duration and Intensity Increments in Melody Tones: A Partial Connection between Music Perception and Performance" (Bruno H. Repp); "Acoustics, Perception, and Production of 'Legato' Articulation on a Digital Piano" (Bruno H. Repp); and "Pedal Timing and Tempo in Expressive Piano Performance: A Preliminary Investigation" (Bruno H. Repp). (RS)

ED 390 082

CS

Haskins Laboratories Status Report on Speech Research

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
A. Dudaiean
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS509076

SR-117/118
JANUARY-JUNE 1994

BEST COPY AVAILABLE

Haskins
Laboratories
Status Report on

Speech Research

SR-117/118
JANUARY-JUNE 1994

NEW HAVEN, CONNECTICUT

Distribution Statement

Editor
Carol A. Fowler

Production
Yvonne Manning-Jones
Fawn Zefang Wang

This publication reports the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications.

Distribution of this document is unlimited. The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.

Correspondence concerning this report should be addressed to the Editor at the address below:

Haskins Laboratories
270 Crown Street
New Haven, Connecticut
06511-6695

Phone: (203) 865-6163 *FAX:* (203) 865-8963
Internet: HASKINS@HASKINS.YALE.EDU

This Report was reproduced on recycled paper

Acknowledgment

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development

Grant HD-01994

National Institute of Mental Health

Grant MH-51230

National Science Foundation

Grant DBS-9112198

National Institute on Deafness and Other Communication Disorders

Grant DC 00121
Grant DC 00183
Grant DC 00403
Grant DC 00016
Grant DC 00594
Grant DC 02151
Grant DC 00865
Grant DC 01147
Grant DC 00044
Grant DC 00825
Grant DC 01247

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Eric Bateson*
Fredericka Bell-Berti*
Shlomo Bentin*
Catherine T. Best*
Susan Brady*
Catherine P. Browman
Dani Byrd
Claudia Carello*
Franklin S. Cooper*
Stephen Crain*
Alice Faber
Laurie B. Feldman*
Janet Fodor*
Anne Fowler*
Carol A. Fowler*
Ram Frost*
Louis Goldstein*
Carol Gracco
Vincent Gracco
Katherine S. Harris*
Leonard Katz*
Rena Arens Krakow*
Andrea G. Levitt*
Alvin M. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist
Georgije Lukatela*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Walter Naito†
Weijia Ni
Patrick W. Nye
Kiyoshi Oshima†
Kenneth Pugh*
Lawrence J. Raphael*
Bruno H. Repp
Hyla Rubin*
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Michael Studdert-Kennedy*
Michael T. Turvey*
Douglas Whalen

Technical Staff

Michael D'Angelo
Vincent Gulisano
Donald Hailey
Yvonne Manning-Jones
William P. Scully
Fawn Zefang Wang
Edward R. Wiley

Administrative Staff

Philip Chagnon
Alice Dadourian
Betty J. DeLise
Lisa Fresa
Joan C. Martinez

Students*

Lawrence Brancazio
Melanie Campbell
Sandra Chiang
Terri Erwin
Douglas Honorof
Pai-Ling Hsiao
Laura Koenig
Simon Levy
Subhobrata Mitra
Mira Peter
Joaquin Romero
Dorothy Ross
Arlyne Russo
Michelle Sancier
Sonya Sheffert
Brenda Stone
Mark Tiede
Qi Wang

*Part-time
†Visiting from University of Tokyo, Japan

Contents

The Universality of Intrinsic F0 of Vowels
D. H. Whalen and Andrea G. Levitt 1

Intrinsic F0 of Vowels in the Babbling of 6-, 9- and 12-month-old French- and English-learning Infants
D. H. Whalen, Andrea G. Levitt, Pai-Ling Hsiao, and Iris Smorodinsky 15

Knowledge from Speech Production Used in Speech Technology: Articulatory Synthesis
Richard S. McGowan 25

Nonsegmental Influences on Velum Movement Patterns: Syllables, Sentences, Stress, and Speaking Rate
Rena A. Krakow 31

Articulatory Organization of Mandibular, Labial, and Velar Movements During Speech
H. Betty Kollia, Vincent L. Gracco, and Katherine S. Harris 49

An Acoustic and Electropalatographic Study of Lexical and Post-lexical Palatalization in American English
Elizabeth C. Zsiga 67

The Discriminability of Nearly Merged Sounds
Alice Faber and Marianna Di Paolo 81

The Role of Fundamental Frequency in Signaling Linguistic Stress and Affect: Evidence for a Dissociation
Gerald W. McRoberts, Michael Studdert-Kennedy, and Donald P. Shankweiler 113

Orthographic Representation and Phonemic Segmentation in Skilled Readers: A Cross-language Comparison
Ilana Ben-Dror, Ram Frost, and Shlomo Bentin 133

Expressive Timing in Schumann's "Träumerei": An Analysis of Performances by Graduate Student Pianists
Bruno H. Repp 141

Quantitative Effects of Global Tempo on Expressive Timing in Music Performance: Some Perceptual Evidence
Bruno H. Repp 161

Detectability of Duration and Intensity Increments in Melody Tones: A Partial Connection between Music Perception and Performance
Bruno H. Repp 173

Acoustics, Perception, and Production of *Legato* Articulation on a Digital Piano
Bruno H. Repp 193

Pedal Timing and Tempo in Expressive Piano Performance: A Preliminary Investigation
Bruno H. Repp 211

Appendix 233

Haskins
Laboratories
Status Report on

Speech Research

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 1-14

The Universality of Intrinsic F0 of Vowels*

D. H. Whalen and Andrea G. Levitt†

The tendency for high vowels such as [i] and [u] to have higher fundamental frequencies (F0s) than low vowels such as [a] has been found in every language so far in which it has been sought. These include 31 languages representing 11 of the world's 29 major language families (as defined by Crystal, 1987). While the size of the intrinsic F0 (IF0) effect varies from study to study, the differences seem to derive from differences in the study design, especially in the number of subjects. The effect appears larger for female speakers when expressed in Hz, but it is, instead, larger for males when the results are expressed in semitones. The size of the language's vowel inventory did not significantly affect the size of IF0. One other universal, though, is that the effect disappears at the low end of a speaker's F0 range. The consistency of intrinsic F0 across languages argues that the effect is truly intrinsic; that is, it is not a deliberate enhancement of the signal but rather a consequence of successfully forming a vowel.

1. INTRODUCTION

Although all spoken human languages make use of sounds produced by the vocal tract, there is a great deal of latitude in the sounds selected. Every language uses vowel sounds, for example, but the number of distinctive vowels ranges from a low of 2 for Margi or Ubykh (Ladefoged & Maddieson, 1990) to a high of 24 (for !Xũ) in the UCLA database of 317 languages (Maddieson, 1984). For languages with the same number of vowels, there is a great range of combinations of vowel qualities selected. These vowels are also the primary locus for lexical suprasegmental distinctions which exist in many languages, such as stress, pitch accent, and tones. Such differences are among the features that make languages distinct.

This project was supported by NIH grants DC-00403 and HD-01994 to Haskins Laboratories. Ken deJong contributed a significant amount of work in extracting the Navaho F0s, for which we are extremely grateful. Julie Lavoie also kindly provided her raw data. We are also grateful to the respondents from the LINGUIST list: Jan-Olaf Svantesson, Jurika Bakran, Ocke Bohn, Michael Jessen, Hartmut Traunmüller, Conrad Ouellon, Florien Koopmans-van Beinum, and Ingegerd Eklund. John Ohala and Harmut Traunmüller were especially helpful in ferreting out ever more references. We thank Carol A. Fowler, Ilse Lehiste, Kiyoshi Honda and John Ohala for helpful comments.

One phonetic feature that has been found to accompany vowels is "intrinsic F0" or "intrinsic pitch" (IF0, from here on). This is the tendency for the high vowels, such as [i] and [u], to have a higher fundamental frequency than the low vowels, such as [a] and [æ]. IF0 was first noticed for German (Meyer, 1896-7) and has since been found in every language that has been examined for it. Our goal in the present paper is to examine the size of the effect in all the languages reported so far to see whether there is any difference that does not seem to be due to the factors of experimental procedure. The features we will study will include not just language but also speaker sex and the size of the language's vowel inventory. The survey will also provide an estimate of the range of variability that can be expected in measuring IF0.

The mechanism behind this effect is the object of considerable debate. We will not give a full history of the different explanations, since such surveys exist elsewhere (DiCristo, Hirst, & Nishinuma, 1979; Shadle, 1985; Silverman, 1987; Sapir, 1989; Fischer-Jørgensen, 1990). All of the explanations surveyed there have assumed that IF0 is an automatic consequence of articulation, most likely the pull of the tongue on the laryngeal system, or acoustics. (Steele (1986) argues that there must be a contribution of subglottal

pressure.) However, more recently there have been proposals that IF0 is simply another deliberate change in F0 that is introduced in the signal to enhance the differences between vowel categories (Diehl & Kluender, 1989b; Diehl & Kluender, 1989a; Diehl, 1991; Kingston, 1993).

The consistency of the effect across the world's languages can provide us with some indication of whether an automatic or deliberate process is more likely. If we find differences among the different languages, it would be likely that the degree of IF0 is another variable that languages choose, just as they choose their inventory or their suprasegmental category. If, on the other hand, there seems to be little change in the size of IF0, then we would expect that, whatever mechanism is responsible, it is truly intrinsic, and occurs as part of the production of vowels in any language. Such a conclusion can only be provisional, of course, since it is (in the statistical sense) the null hypothesis. Nonetheless, a failure to observe a difference in the magnitude of the IF0 effect would pose a challenge for any enhancement explanation of the phenomenon.

2. The Published Results

Our data come from various published sources, listed along with the IF0 values in Table 1. These include all of the articles published in journals, proceedings and working papers that we were able to locate. In many cases, IF0 is not directly assessed in the work cited, but the numbers from which IF0 could be calculated were given. Table 2 lists information about the languages, including the size of the vowel inventory.

Table 1. *Values for intrinsic F0 from published sources. All F0 values are in Hz. Values in italics were not used in the statistical test. Those studies that lacked values for [u] were also not used for the statistical test, but are included in Figure 1. If results were combined for males and females, the sex is listed as "B." In the "Context," "V" stands for vowel and "S" for sentence.*

Language, Source	#, Sex of S's	Context	F0 for [u]	F0 for [i]	F0 for [a]
English					
Crandall, 1925	4 M	Isolated V's	140	136	113
	4 F	Isolated V's	270	252	234
Taylor, 1933	8 M	Isolated real words	152	149	132
	9 F	Isolated real words	323	320	298
Black, 1949	16 M	Isolated CVC	153	146	133
Peterson & Barney, 1952	33 M	Isolated hVd	141	136	124
	28 F	Isolated hVd	231	235	212
House & Fairbanks, 1953	10 ?	?	130	128	118
Lehiste & Peterson, 1961	5 M	?	124	129	120
Peterson, 1961	4 M	Isolated V	128	124	119
	3 F	Isolated V	253	250	212
Atkinson, 1973	5 M	Sentences	128	132	114
Fox, 1982	8 F	Isolated hVd	240	242	234
	8 M	Isolated hVd	145	148	140
Shadle, 1985	2 F	Sentence, first 3 positions	225	225	215
		Sentence, last position	*175*	*174*	*175*
	2 M	Sentence, first 3 positions	116	115	108
		Sentence, last position	*93*	*93*	*90*
Zawadzski & Gilbert, 1989	5 M	?	208	206	188
Nittrouer, McGowan, Milenkovic, & Beehler, 1990	4 M	Nonsense CV in S	139	138	132
	4 F	Nonsense CV in S	215	210	197
Higgins et al., 1994	11 M	Isolated [pa] or [pi]	---	125	120
		Sentences w/ [pap] or [pip]	---	*126*	*121*
	10 F	Isolated [pa] or [pi]	---	206	196
		Sentences w/ [pap] or [pip]	---	*212*	*194*
Hillenbrand et al., 1995	45 M	Isolated hVd	143	138	123
	48 F	Isolated hVd	235	227	215
Dutch					
Koopmans-van Beinum, 1980	2 M	Monosyllabic words	153	149	126
	2 F	Monosyllabic words	217	229	205
van Son, 1993	1 M	Running text (+accent only)	197	193	186

Table 1. *continued*

German					
Meyer, 1896-7	1 M	Isolated V's	121	108	87
Mohr, 1971	1 M	Initial V	122	121	116
Neweklowsky, 1975	2 M	Isolated pVp (short V)	139	137	124
(all are Viennese)	1 F	Isolated pVp (short V)	219	222	210
Antoniadis & Strube, 1981	3 M	CVCə (long V)	132	131	125
		CVCə (short V)	*135*	*133*	*129*
Traunmüller, 1982	12 M	Nonsense CV in S	126	133	105
Möbius, Zimmermann, & Hess, 1987	1 M	Real words in carrier S (long V)	116	123	107
Iivonen, 1989	5 M	Isolated words (long V)	95	90	82
Iivonen, 1989 (Viennese)	5 M	Isolated words (long V)	140	138	106
Fischer-Jørgensen, 1990	5 M	dVdə in carrier S (long V)	---	125	112
	1 F	dVdə in carrier S (long V)	---	184	167
Danish					
Reinholt Petersen, 1978	2 F	CVCV:CV in S (middle syll)	220	212	182
	3 M	CVCV:CV in S (middle syll)	122	120	104
Swedish					
Fant, 1959	7 M	Isolated Vs	127	128	124
	7 F	Isolated Vs	222	218	215
French					
Rossi & Autesserre, 1981	4 M	Isolated V's	142	140	131
DiCristo, 1982	1 F	?	243	235	226
	3 M	?	134	132	124
Lavoie, 1994 (Quebecois)	2 M	Short phrases	198	199	185
	2 F	Short phrases	300	289	266
Italian					
Pettorino, 1987	1 M?	Isolated real words	158	159	143
Greek					
Samaras, 1972	2 M	CVC	148	149	137
Russian					
Mohr, 1971	1 M	Initial V	127	126	121
Bolla, 1981	1 M	Isolated real words	135	136	116
Polish					
Steffen-Batóg, 1970	6 M	Isolated VCV	153	153	150
Serbo-Croatian					
Ivic & Lehiste, 1965	12 B	Words in initial and medial			
		position in S, beg. of syll	183	217	194
		Final S position, beg. of syll	*129*	*138*	*159*
Standard Croatian					
Bakran & Stamenković, 1990	3 M	?	108	110	98
	3 F	?	177	197	165
Lithuanian					
Pakerys, 1982	7 M?	Isolated words - stressed	149	152	139
		- unstressed	*128*	*126*	*125*
Hindi					
Schiefer, 1987	1 F	Isolated real words	252	238	205
Gujarati					
Dave, 1967	3 M	Isolated Vs, words, phrases	136	145	128
Finnish					
Vilkman, et al., 1989	1 M	Isolated Vs	119	118	114
Hungarian					
Tamas, 1976	? M	stressed vowels	137	144	120
		unstressed vowels	*103*	*90*	*85*
	? F	stressed vowels	231	235	205
		unstressed vowels	*195*	*185*	*170*
Korean					
Han & Weitzman, 1967	1 M	Nonsense CV	185	183	166
	1 F	Nonsense CV, CVC	329	316	317
Kim, 1968	? F	?	279	277	268

Table 1. *continued*

Japanese					
Homma, 1973	1 F	Words?	342	350	328
Nishinuma, 1979	5 M	2nd syll of 4 syll words	148	142	136
"Chinese"					
Mohr, 1971	1 M	Initial V	148	150	147
Mandarin					
Shi & Zhang, 1987	5 M	Tone 1 (= high F0)	181	175	154
		Beg. of Tone 2 (= low F0)	*117*	*118*	*111*
		End of Tone 2 (= high F0)	*168*	*167*	*151*
		Beg. of Tone 3 (= mid F0)	*112*	*113*	*108*
		End of Tone 3 (= low F0)	*90*	*89*	*83*
		Beg. of Tone 4 (= high F0)	*206*	*197*	*175*
		End of Tone 4 (= low F0)	*105*	*97*	*97*
Shi & Zhang, 1987	5 F	Tone 1 (= high F0)	307	297	276
		Beg. of Tone 2 (= low F0)	*209*	*205*	*198*
		End of Tone 2 (= high F0)	*289*	*265*	*255*
		Beg. of Tone 3 (= mid F0)	*218*	*219*	*227*
		End of Tone 3 (= low F0)	*172*	*169*	*171*
		Beg. of Tone 4 (= high F0)	*335*	*312*	*302*
		End of Tone 4 (= low F0)	*184*	*180*	*187*
Taiwanese					
Zee, 1980	3 M	Sentences, High Tone	157	157	143
		Sentences, Low Tone	*100*	*99*	*102*
Shanghai					
King, Ramming, Schiefer, & Tillmann, 1987	1 M	Isolated words, Tone 2, end	156	163	142
Kammu					
Svantesson, 1988	1 M	Words in S: High tone	150	148	129
		: Low tone	*115*	*116*	*104*
Paraok					
Svantesson, 1993	1 M	Isolated words	159	156	142
	1 F	Isolated words	293	305	250
Vietnamese					
Han, 1969	1 F	Vs in carrier S: High tone	308	309	288
		End of falling tone	*185*	*189*	*194*
	1 M	Vs in carrier S: High tone	143	141	136
		End of falling tone	*116*	*118*	*107*
Thai					
Gandour & Maddieson, 1976	1 M	Isolated nonsense CV, middle of high tone	148	147	138
Malagasy					
Rakotofiringa, 1968	1? M?	?	124	123	122
Rakotofiringa, 1982	18 M	Isolated words	122	119	116
	11 F	Isolated words	238	233	225
Yoruba					
Hombert, 1977	2 ?	Isolated V, high tone	177	179	170
		Isolated V, mid tone	*152*	*153*	*147*
		Isolated V, low tone	*121*	*119*	*120*
Itsekiri					
Ladefoged, 1968	1 M	?	(reports a 5 Hz difference)		
Hausa					
Pilszczikowa-Chodak, 1972	1 M	Isolated words	122	141	117
Navaho					
deJong & McDonough, 1993	4 F	Words, High tone	---	209	188
		Words, Low tone	---	*181*	*180*
	2 M	Words, High tone	---	133	138
		Words, Low tone	---	*129*	*129*

see footnote 1

Table 2. *Information about the languages. Family affiliation is taken from Crystal (1987). The number of vowels represents the number of distinctive vowel qualities in the articulatory space (tongue height, front/back, rounding). Thus vowels that differed distinctively in duration but had the same quality were considered to represent one vowel, not two. Distinctive nasality was also ignored. See text for the rationale for this decision.*

Language	Family	Sub-Family	# of Vs
English	Indo-European	W.Germanic	12
Dutch	Indo-European	W.Germanic	12
German	Indo-European	Germanic	14
Danish	Indo-European	N.Germanic	12
Swedish	Indo-European	N.Germanic	18
French	Indo-European	Italic	12
Italian	Indo-European	Italic	7
Greek	Indo-European	Greek	5
Russian	Indo-European	Slavic	5
Polish	Indo-European	Slavic	6
Serbo-Croatian	Indo-European	Slavic	5
Standard Croatian	Indo-European	Slavic	5
Lithuanian	Indo-European	Balto-Slavic	11
Hindi	Indo-European	Indo-Iranian	14
Gujarati	Indo-European	Indo-Iranian	8
Finnish	Uralic	Finno-Ugric	16
Hungarian	Uralic	Finno-Ugric	10
Korean	Korean		18
Japanese	Japanese		5
"Chinese"	Sino-Tibetan	Sinitic	5?
Mandarin	Sino-Tibetan	Sinitic	5?
Taiwanese	Sino-Tibetan	Sinitic	6
Shanghai	Sino-Tibetan	Sinitic	12
Kammu	Austro-Asiatic	Mon-Khmer	10
Paraok	Austro-Asiatic	Mon-Khmer	9
Vietnamese	Austro-Asiatic	Mon-Khmer	11
Thai	Tai	S.-Western	9
Malagasy	Austronesian	Western	4
Yoruba	Niger-Congo	Kwa	7
Itsekiri	Niger-Congo	Kwa	?
Hausa	Afro-Asiatic	Chadic	5
Navaho	Na-Dené	Athabaskan	4

Some of the descriptions are incomplete, and so question marks appear in some places. In some cases, it was difficult to find the appropriate measures. For example, the Serbo-Croatian results (Ivic & Lehiste, 1965) have been cited a number of times as showing IF0, and yet there is no condition in which each of the vowels appears with the same pitch accent. We have tried to minimize the effect of accent for this study of Serbo-Croatian by averaging the values at the beginning of the vowel, a point at which the F0 excursion for the pitch accent should not be as advanced as later. Only four words were measured (brâtu, grâdu, Mári, sèlu), which is a smaller set than would be best. Any perturbation due to consonant voicing should be minimal, since voicing is different for only one of the syllables. These Serbo-Croatian values would be more comparable to the others if we had instances of the different vowels with the same accent. Because of these difficulties, these numbers are not included in the statistical tests reported below. The numbers for Navaho (deJong & McDonough, 1993) are not represented directly in that paper and have been supplied by the authors. Similarly, the F0 values for Quebecois French were not in the published version and were supplied by the author (Lavoie, 1994). The "Chinese" of Mohr (1971) is presumably Mandarin (and is so treated in the analysis of variance), but has been listed as "Chinese" in the table.

We restricted our analysis to the vowels [i], [u], and [a] (or [ɑ]). These include the two dimensions

that have been examined most, vowel height and front/back. They are present in about 80% of the world's languages (Maddieson, 1984), and occur in almost all the languages examined: Only Navaho lacked one of the vowels, namely [u]. However, the high back vowel in the Tokyo dialect of Japanese (the one studied in both references here) is an [ɯ] rather than [u]. There were two other studies that lacked [u] values, simply because they were not measured. One was for English (Higgins, Netsell, & Schulte, 1994) and one for German (Fischer-Jørgensen, 1990). Our focus on these three vowels is not intended to deny that there is gradation between high and low or that the other dimensions of rounding, tense/lax, nasalization, advanced tongue root, etc., might not play a part in IF0. The selection was made solely to equate the language samples as far as was possible.

The numbers come from published studies, and usually there is no individual data given, so it is not possible to perform ordinary statistics on these numbers. We performed an ANOVA on the averaged results for most of the numbers in Table 1. Italicized numbers were not used, since they represent different phonetic environments. They are included as comparisons to the main environment studied in the statistical analysis. We also excluded from the statistical analysis the studies which did not have measurements for [u]. Each study was weighted by the number of subjects measured, so that they contributed to the means in proportion to the amount of data represented. We performed separate tests for a number of factors: vowel ([i] versus [u] versus [a]), front/back ([i] versus [u]), sex, language, English versus the other languages, and languages compared by vowel inventory size. Separate analyses were conducted because there were not enough studies to have anything like equal cell sizes if we included more than one factor. We also performed each test twice, once on the Hz values and once on the values expressed in semitones. The use of semitones helps equate the differences found for the males and females. Since the semitone scale relates one frequency to another, we needed to have a reference frequency. We chose 1 Hz, so that the denominator would fall out, and the simple formula:

$$\text{semitone} = 1/\log(2) * 12 * \log(\text{Hz})$$

could be used. The values that we were able to use are somewhat less than optimal, since they are semitone transforms from the mean expressed in Hz, rather than a mean of the values expressed as semitones. But since the majority of studies did not give individual values, this was our only choice.

Our first analysis looked at a single factor with three levels, namely, a vowel effect comparing [u]/[i]/[a]. The vowel effect was highly significant ($F(2,150) = 246.67$; $p < .0001$). Overall, the means were 177.4 Hz for [u], 174.9 Hz for [i], and 160.9 Hz for [a]. This translates into a 15.3 Hz difference for IF0 across all the languages. The semitone analysis was also highly significant ($F(2,150) = 231.51$; $p < .0001$). Those means were 88.90, 88.66, and 87.13 semitones for [u], [i] and [a] respectively. That difference is 1.65 semitones. As expected, then, IF0 is a highly significant effect.

Front/back (comparing just [i] and [u]) was not a significant factor, either in the Hz analysis ($F(1,76) = 2.48$, n.s.) or the semitone analysis ($F(1,76) = 3.16$, $p < .10$). As can be seen from the means in the previous paragraph, there is a 2.5 Hz (0.24 semitone) difference between the front and back vowels across all the languages. Inspection of Table 1 will show that some languages seem to have large differences between the front and back vowels. For example, English [u] is reliably higher in F0 than [i] as measured across the studies. The 184.7 Hz of [u] is reliably (3.6 Hz) higher than the 181.1 of [i] ($F(1,20) = 12.73$, $p < .01$) as is the .36 semitone difference in the semitone analysis ($F(1,20) = 15.19$, $p < .001$). While this may indicate that there is a genuine difference in the front and back vowels for English, this analysis would be more reliable if done with the individual measurements rather than the means across experiments. The amount of variability that is present within experiments is not carried across when only the means are used. A comparison case is German, for which we have the next largest data set. In Table 1, six of the nine entries for German also show a higher value for [u] than [i], but one of the studies that has higher values for [i] than [u] also had a large number of subjects (Traunmüller, 1982), and so the overall effect is not significant and, in fact, shows [i] being 0.6 Hz higher than [u]. Again, it does not seem impossible that a language could adopt an articulation for one vowel or another that would modify its IF0, but there is no indication of a universal difference between [i] and [u]. It is possible that the English differences are real, and therefore that some languages have effects that other languages lack, indeed, that other

languages have in the other direction. The effect in English is one fourth the size of the height effect and will consequently be harder to verify. It is also likely that some of the apparent differences are simply the result of the inevitable sampling error.

For the remainder of the analyses, we will collapse across [i] and [u] and subtract the value for [a], giving us a simple test for the significance of the difference in height. The test for the means (i.e., whether there was an IF0 effect) was always significant, and so we will not report the numbers. In this way, the factors that we examine from here on out will directly indicate whether there was an effect of that factor on the size of the IF0 difference.

For the analysis by sex, we excluded the three studies that did not separate out the two sexes (House & Fairbanks, 1953; Ivic & Lehiste, 1965; Hombert, 1977). The overall IF0 effects were 13.9 Hz for the males and 15.4 for the females. This difference in magnitude was significant ($F(1,72) = 5.87$, $p < .05$). It would appear from these numbers that women have a larger effect than men. But they also have a higher overall F0, so the greater magnitude of their effect represents, in fact, a smaller percent increase. This fact is accounted for in the semitone analysis. That analysis, though, also shows an effect ($F(1,72) = 6.53$, $p < .05$), but now the difference is larger for the males than the females (1.84 semitones for the males and 1.34 for the females). It appears that there is a difference in the size of IF0 by the sex of the speaker. As with the front/back difference, though, this indication must be treated with caution since it was, of necessity, based on the means across experiments and not, as would be optimal, on the individual results. Thus the lack of a contribution of individual variability may have made a small, random difference appear significant. This aspect of the data will require further elaboration before we can decide whether there is anything that needs explaining. However, the study with the largest number of subjects (Hillenbrand et al., 1995) had a sizable difference between males and females, so it may simply be a factor that requires substantial evidence before it is apparent. While it is conceivable that the lowering of the male larynx during puberty could be a factor in such a difference, no explanation is immediately obvious.

The analysis by language did not reveal an IF0 difference for any one language that was statistically different from the others ($F(27,48) = 1.27$, n.s. for the Hz analysis; $F(27,48) = 1.61$, n.s. for the semitone analysis). Since approximately half of our results came from English, it was of interest to see whether the English results patterned with the others. We did this by classifying each IF0 measurement as being either English or non-English. This also showed no difference ($F(1,74) = 2.39$, n.s. for the Hz analysis; $F(1,74) < 1$, n.s. for the semitone analysis). Thus there is no evidence that English stands out from the others in terms of the size of the IF0 effect. This test, besides its weakness as accepting the null hypothesis, must also be treated with caution because so many of the languages were poorly represented. If there were just one language that showed an effect of a different size, it would be very hard to detect without measuring many more subjects. The one language (English) that we do have many measurements for, though, seems to be typical.

Finally, we separated the languages by the size of their vowel inventory. Since we have such a large amount of data on English, the 12-vowel systems are automatically overrepresented. We thus classed them by themselves. For the rest, we divided them into small (4-5) medium (6-11) and large (13 or more) vowel inventories. The inventory size is taken as the number of monophthongal distinctive vowels that differ in quality. In many cases, these same vowels could also differ in length or nasalization, but that was not considered to make a different vowel for this analysis. If length makes a distinction without changing quality, then the long and short were counted as one vowel, while long and short vowels that did differ (such those in German) were counted as two vowels. This form of counting vowels was adopted under the assumption that any enhancement that might be introduced deliberately would be likely to depend on how crowded the vowel space is. Even if such separable dimensions as length and nasalization are distinctively for a language, they should not influence any use of F0 in perception, since F0 is posited to affect primarily the height dimension.

The size of the IF0 effect was 11.6, 12.6, 16.3 and 15.1 Hz for the small, medium, 12-vowel and large inventories respectively. These did not differ significantly ($F(3,72) = 1.56$, n.s.). The size of the IF0 effect in semitones was 1.17, 1.33, 1.70 and 1.64 for the small, medium, 12-vowel and large inventories respectively. These also did not differ significantly ($F(3,72) = 2.47$, n.s.).

Again, we need a more balanced set of results before we can be sure that these statistics do not hide a small effect, but the current indication is that vowel inventory size also does not affect the size of IF0.

For some of the 10 tone languages in the survey, there is an additional observation that can be made: The low tones fail to show IF0. While this is true of all but one of the tone languages in Table 1, perhaps the most striking case is the Tone 4 data of Shi and Zhang (1987). There, within one syllable, the IF0 effect was present at the onset (the high portion of the F0 contour) and then absent at the offset (the low frequency). Such results match up with the findings of Shadle (1985) and Ladd and Silverman (1984) for spontaneous speech (for English and German respectively). In fact, Ladd and Silverman claim that "low pitch, rather than low-stress or phrase-final position per se, is the relevant factor" (1984:36). Of course, this statement is slightly misleading, since what is important is the relative frequency within a speaker's range rather than absolute F0. Still, the main point is borne out by the tone languages, that in the lower part of the frequency range, IF0 disappears. The one case where it does not is Kammu. The high tone has an effect of 20 Hz while the low has an effect of 12 Hz. This is a fairly substantial difference, but the 12 Hz on the low tone is certainly within the range of values found for other studies with just one speaker. Although this language may constitute an exception to the disappearance of IF0 at the lower part of the F0 range, it is also possible that the "low" tone is simply low in relation to the high tone, and that a mid-range pitch is used for it. In that case, we would expect to find an IF0 effect. Further study is needed before deciding this issue, but it is interesting to compare the Kammu results with the Yoruba in Table 1. Yoruba has three tones, high, mid and low. The IF0 effect is 8, 6 and 0 Hz for those, respectively. So the existence of an IF0 effect for a "lower" (i.e., mid) tone is not out of the question.

The range of variability shown by these studies appears rather large, but there is a tendency, noted above, for the larger studies to approach more closely the mean across studies (without being weighted by number of subjects). Such an outcome is to be expected if the measurements come from a single distribution, but it is not a necessary outcome if some languages differ from the rest. The measured IF0 does differ across studies, but this appears to be due to typical distributional factors. There are three sources of variability in the published results. The first source is the distribution of IF0 values for different languages. Since speech is a natural system, we have to expect some variation in even its most consistent aspects. If we analyzed a large number of speakers for each language (and languages truly do not differ), the distribution of IF0 values would be extremely tight. If languages were selecting from a small range of choices (e.g., either the typical IF0 or none at all), then we might find a bimodal distribution. The second source of a distribution is in the individual IF0 values. If we looked closely at values for individual speakers within a language, we would expect to see somewhat larger variability than for the languages themselves, since it is easier for an individual to be extreme than for a whole language community. These first two are sources of variability in the real world. A final source of variability is in our measures of the world. These will also be distributed around the true mean, since any measurement is inherently fallible. This measure will also tend to form tighter distributions as the number of subjects increases.

As a way of judging the combined effects of these three distributions, we have plotted in Figure 1 the difference between the value for [a] and the average for [i] and [u] from Table 1. These include all the lines from Table 1 that were used in the analysis of variance, along with the values from the studies lacking an [u]. The top panel shows the results for the Hz analysis, and the bottom, for the semitone analysis. The number of subjects in each study is shown on the y-axis. We have plotted the data this way rather than doing a more typical histogram because there are so many gaps in our distribution, in terms of the number of subjects per study. While we certainly do not have a large enough sample to show that the values are distributed normally, there is a tendency for the larger studies to show values close to the overall mean difference of 15.3 Hz . The semitone analysis shows an even tighter distribution, with both ends of the distribution pulling in towards the mean of 1.65 semitones. Again, only a much larger set of results for all of the languages in this set would allow us to test for this tendency statistically, but the distributions shown in Figure 1 are consistent with an absence of any difference across languages. It is also easy to find where a new set of results would fit in to see whether it conforms to the current pattern.

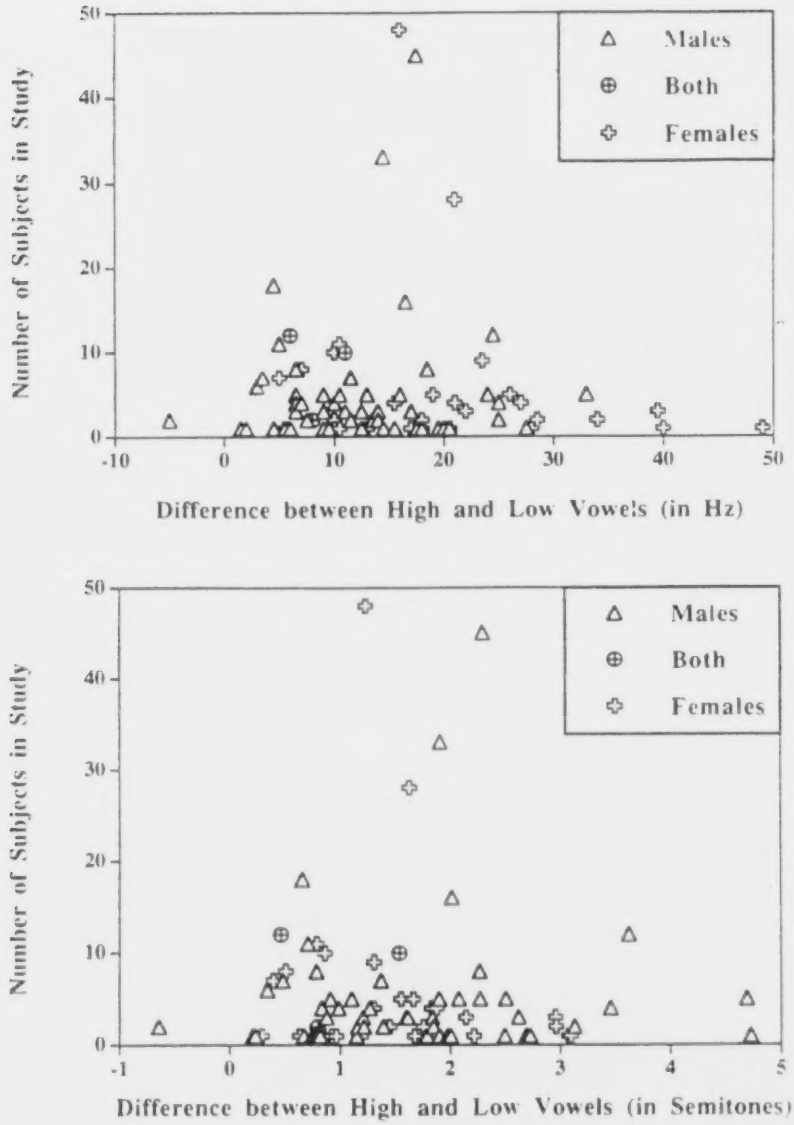


Figure 1. **Distribution of the differences between the two high vowels and the low vowel for the studies in Table 1. The top panel represents the differences between the values expressed in Hz, and the lower panel, between the values expressed as semitones.**

3. IF0 as Enhancement?

Diehl and colleagues (Diehl & Kluender, 1989b; Diehl & Kluender, 1989a; Diehl, 1991) have claimed that IF0 is an enhancement of the speech signal. Enhancements, in their view, are deliberate attempts by the speaker to make the auditory aspects of speech more salient for the hearer. Perception of vowel height is claimed to be influenced by the difference between F1 and F0. IF0, then, enhances this difference, since a high vowel will have a low F1. If this low F1 is accompanied by a high F0, the difference will be even smaller than before, and thus even more distinct from the large difference for low vowels.

One aspect of other proposed enhancements is that they are optional. For example, the typical vowel systems of the world combine rounding

BEST COPY AVAILABLE

with back vowels, to increase the proximity of F2 and F1 (Lindblom, 1986). Yet there are languages, such as the Iroquoian languages, which avoid the use of lip rounding (e.g., Oneida: Lounsbury, 1953:27). If IF0 is truly universal and unmodified, it is less likely to be under voluntary control. Of course, even if we had data on every language of the world, it still might be the case that they all happened to choose to enhance the vowels this way, and that the next human language to evolve would dispense with IF0. In the absence of even one example of a language doing without, however, the enhancement view is tenuous.

The magnitude of the shift in vowel identity that we might expect from the IF0 differences is rather small. In a survey of such studies, Nearey (1989) found the following size of shifts in the frequency of F1 for an octave shift in F0: 16%, 6%, 14%, 21%, and 26-30% (see his page 2093). As a rough estimate, then, we might expect a 17% shift in effective F1 for an octave shift in F0. The IF0 effect, however, is much smaller than an octave, approximately one seventh that size (around 1.7 semitones), which would lead us to expect about a 2.4% shift in the effective F1. This value is below the difference limen (of 3-5%) for changes in single formants of vowels (Flanagan, 1955). It is difficult to see why such a complicated change in vowel articulation would be deliberately introduced for such a modest reward, especially for those languages with 4-5 vowels. The enhancement account would seem more likely if there were an effect of vowel inventory size on the size of IF0, so that more vowels would lead to a greater use of IF0. The present results showed no such effect, however.

It is also hard to see why tone languages would want to use F0, which is critical for tone, to enhance vowel identification, especially in a system like Mandarin's in which the vowel system is quite simple (5 vowels). The tones occur with every vowel, and therefore the tonal differences of over 100 Hz appear with each vowel, yet the 5-10 Hz IF0 effect remains. Perhaps a parsing of the F0 into its tone and IF0 components would make this work, but such an effect is indicative of a perceptual system operating on the signal, not the low-level auditory effect that the enhancement theory presupposes. It seems more likely that Mandarin and other tone languages exhibit IF0 because it is a natural consequence of producing vowels of different heights: Even when speakers change the F0 of a vowel (within the upper part of the F0 range, at least), there is a contribution of IF0 present.

One other aspect of IF0 that appears to favor a deliberate component is the fact that IF0 appears in some laryngectomized patients who use a flap over the esophagus to voice their sounds (Gandour & Weinberg, 1980; Pettorino, 1987), but there is a problem with claiming that such evidence means that IF0 in normal speech is deliberately produced. Even if we assume that these speakers are introducing IF0 deliberately, we would not know anything more about the normal case: The esophageal speakers could simply be recreating what used to come to them naturally. If they monitor their own productions and find that their [i]'s, for example, are consistently too low in frequency to sound right, then they may deliberately raise the F0. The IF0 of esophageal vowels does not indicate that normal IF0 is deliberate.

If it is important that the vowel judgment be enhanced, it is not clear why speakers would choose not to exhibit IF0 in the lower portion of their frequency range. While it is certainly the case for English that words that end up in the low range are unstressed and therefore perhaps dispensable, the low tones of tone languages can occur with any word, perhaps the most important one in the sentence. Kingston (1993) has claimed that the lack of an IF0 effect in the lower frequency ranges constitutes evidence in favor of the enhancement account, but only because the automatic account does not have a ready explanation for it. Conceivably, the arrangement of the larynx makes it very difficult to introduce these differences at low frequencies, and thus the gain from enhancement is judged not to be worth the physiological cost. We will propose below that there is, indeed, something physiologically different about the lower portion of the F0 range. Certainly, other muscles become involved in the lower frequency range (Hallé, 1994). It would be easier to judge whether this lack of an effect in the low frequencies is also automatic or just a cost decision if we could make quantitative predictions about just how much energy is involved. Unfortunately, to make such estimates would require that we would know enough about the laryngeal system that we would know the answer already--we would already know whether IF0 was a necessary consequence of the way the larynx is employed in speech.

The most direct piece of evidence for the enhancement theory is the fact that cricothyroid

(CT) activity increases for high vowels (Autesserre, Roubeau, DiCristo, Chevrie-Muller, Hirst, Lacau, & Maton, 1987; Vilkman, Aaltonen, Raimo, Arajärvi, & Oksanen, 1989; Honda & Fujimura, 1991). Increasing CT activity generally increases F0. Thus it appears that these subjects were intentionally increasing F0 for the high vowels. However, the laryngeal system is quite complex, and the increase in CT activity may not be exclusively associated with higher F0. Vilkman et al. (1989), in fact, consider the intentional explanation and reject it in favor of one in which the CT activity is increased to "avoid opening of the cricothyroid visor during increased vertical pull in the laryngeal region" (page 202). We are currently developing an EMG study to further examine this issue.

In sum, the proposal that IF0 is a deliberate enhancement of the speech signal seems somewhat tenuous. Even when F0 is used extensively for other purposes and vowels occur with the whole range of F0s, as in tone languages, IF0 persists. It would seem that only a perceptual parsing of the F0 signal could disentangle these effects, and this stage is later than the proposed enhancement effect, which is meant to be a low-level one. IF0 disappears at the lower range of F0s, a fact that is difficult to encompass in an enhancement account. Other enhancements of speech occasionally fail to be adopted by one language or another, while IF0 seems to be universal. All of these facts together make it appear that IF0 is truly "intrinsic" and not a deliberate enhancement of the speech signal.

4. Toward the Source of IF0

Our survey has provided some useful limitations on what a theory of IF0 must accommodate. First is its universality. We found no evidence of unusual languages in our survey, and we have examined data from languages belonging to more than a third of the world's language families. Any theory that explains IF0 must deal with the difference between the effect in the non-low portion of the F0 range and the low. In the low region, IF0 disappears. We believe that this is the difference between the changes in F0 that can be accomplished via subglottal air pressure and CT activity versus the changes that need active lowering via the strap muscles. If the strap muscles completely counteract the effect of the tongue on the hyoid bone, then we would expect for low F0 vowels to have no IF0. Similar speculations appear in Pettorino (1987). Much more work remains to be done before this theory can be fully tested.

While our survey has not revealed any language that seems to have an especially exaggerated IF0 effect, there are some articulatory sources for an exaggerated effect. In one experimental paradigm, if a speaker's jaw is fixed to a more open position than is normal for a vowel, then she will compensate with an exaggerated tongue movement. A stronger pull of the tongue should result in a larger IF0 difference. This is what was found in an experiment by Ohala and Eukel (1987). While this experiment does not fully determine the mechanism involved, it does indicate that the action of the tongue is critically involved.

Another instance of an exaggerated effect is found in the speech of the deaf. Bush (1981) found larger IF0 differences for deaf children compared with normal controls. Clearly, the source of this difference cannot be due to an auditorily based enhancement of the height difference. Part of the effect may have been the use of a higher F0 (which may by itself increase IF0; see above). But there seems to be an element of exaggerated articulation that by itself also increases the IF0 effect. Perkell, Lane, Svirsky and Webster (1992) also found an exaggerated difference for subjects prior to receiving a cochlear implant, and this exaggeration disappeared after the implantation. The finding has not been universal, however. Lane and Webster (1991), studying some of the same subjects as Perkell et al. (1992), found no evidence of IF0 at all in their deaf subjects. Lane and Webster measured productions of a full text (the Rainbow Passage), while Perkell et al. examined keywords in a carrier sentence. Perhaps the greater variability in the text production washed out any vowel effect that there might have been. But it does seem that certain forms of production can affect the IF0.

There does not appear to be a developmental trend in IF0. In another paper (Whalen, Levitt, Hsiao, & Smorodinsky, 1995), we show that even infants babbling at 6 months show IF0. While it is true that, since every language has IF0, infants could be imitating IF0, it is hard to see how they could do this intentionally, since the infants do not have any vowel categories to enhance. Similarly, the data of Peterson and Barney (1952), Peterson (1961), Sorenson (1989), Glaze, Bless and Susser (1990) and Hillenbrand, Getty, Clark and Wheeler (1995) also do not show

any change in the size of the effect in the range from 5 to 11 years of age. If enhancements are something to be learned after the major components of a category are mastered, then the case for enhancement status of IF0 is again weakened.

The difference between males and females is in need of verification and an explanation. Since many of the studies examined here involved only one sex or the other, it is difficult to fully accept the difference that appeared. If it holds up in further studies, it seems that the changes in male voices at puberty is likely to be responsible.

IF0 appears to be universal. Our sampling of 31 languages, while far short of the 6000 or so total languages, covers a fairly wide range of families (11 of 29) and language types (tone, pitch accent, and stress). If languages that have no IF0 constitute some small percentage of languages, then we might yet discover a language that makes no use of IF0 at all. Although such propositions can never be fully tested, we feel that the results so far justify the assumption that IF0 is universal. We can, then, for the moment, conclude that IF0 is not a deliberate enhancement of the signal but rather a direct result of vowel articulation.

REFERENCES

Antoniadis, Z., & Strube, H. W. (1981). Untersuchungen zum <<intrinsic pitch>> deutscher Vokale. *Phonetica, 38,* 277-290.

Atkinson, J. (1973). *Aspects of intonation in speech: Implications from an experimental study of fundamental frequency.* Unpublished doctoral dissertation, University of Connecticut, Storrs.

Autesserre, D., Roubeau, R., DiCristo, A., Chevrie-Muller, C., Hirst, D., Lacau, J., & Maton, B. (1987). Contribution du cricothyroidien et des muscles sous-hyoidiens aux variations de la frequence fondamentale en français: Approche electromyographique. In *Proceedings XIth International Congress of Phonetic Science,* 3 (pp. 35-38). Tallinn, Estonia: Academy of Sciences of the Estonian SSR.

Bakran, J., & Stamenkovic, M. (1990). Inherentna frekvencija larngalnog tona u Hrvatskom standardnom jeziku. *Govor, 7,* 1-20.

Black, J. W. (1949). Natural frequency, duration and intensity of vowels in reading. *Journal of Speech and Hearing Disorders, 14,* 216-221.

Bolla, K. (1981). *A conspectus of Russian speech sounds.* Köln: Böhlau.

Bush, M. (1981). *Vowel articulation and laryngeal control in the speech of the deaf.* Unpublished doctoral dissertation, M.I.T.

Crandall, I. B. (1925). The sounds of speech. *Bell System Technical Journal, 4,* 586-626.

Crystal, D. (1987). *The Cambridge encyclopedia of language.* Cambridge: Cambridge University Press.

Dave, R. (1967). A formant analysis of the clear, nasalized, and murmured vowels in Gujarati. *Indian Linguistics, 28,* 1-30.

deJong, K., & McDonough, J. (1993). Tone and tonogenesis in Navajo. *UCLA Working Papers in Phonetics, 84,* 165-182.

DiCristo, A. (1982). *Prolegomenes a l'etude de l'intonation: Micromelodie.* Paris: Editions du Centre National de La Recherche Scientifique.

DiCristo, A., Hirst, D. J., & Nishinuma, Y. (1979). L'estimation de la F0 intrinsèque des voyelles: Etude comparative. *Travaux de L'Institut de Phonétique D'Aix-en-Provence, 6,* 149-176.

Diehl, R. L. (1991). The role of phonetics within the study of language. *Phonetica, 48,* 120-134.

Diehl, R. L., & Kluender, K. R. (1989a). On the objects of speech perception. *Ecological Psychology, 1,* 121-144.

Diehl, R. L., & Kluender, K. R. (1989b). Reply to commentators. *Ecological Psychology, 1,* 195-225.

Fant, G. (1959). Acoustic analysis and synthesis of speech with applications to Swedish. *Ericsson Technics, 1,* 3-108.

Fischer-Jørgensen, E. (1990). Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica, 47,* 99-140.

Flanagan, J. L. (1955). A difference limen for vowel formant frequency. *Journal of the Acoustical Society of America, 27,* 613-617.

Fox, R. A. (1982). Individual variation in the perception of vowels: implications for a perception-production link. *Phonetica, 39,* 1-22.

Gandour, J., & Maddieson, I. (1976). Measuring larynx movement in Standard Thai using the cricothyrometer. *Phonetica, 33,* 241-267.

Gandour, J., & Weinberg, B. (1980). On the relationship between vowel height and fundamental frequency: Evidence from esophageal speech. *Phonetica, 37,* 344-354.

Glaze, L. E., Bless, D. M., & Susser, R. D. (1990). Acoustic analysis of vowel and loudness differences in children's voice. *Journal of Voice, 4,* 37-44.

Hallé, P. A. (1994). Evidence for tone-specific activity of the sternohyoid muscle in modern standard Chinese. *Language and Speech, 37,* 103-123.

Han, M. S. (1969). *Vietnamese tones.* Studies in the Phonology of Asian Languages, vol. 8. Los Angeles: University of Southern California.

Han, M. S., & Weitzman, R. S. (1967). *Acoustic features in the manner-differentiation of Korean stop consonants.* Studies in the Phonology of Asian Languages, vol. 5. Los Angeles: University of Southern California.

Higgins, M. B., Netsell, R., & Schulte, L. (1994). Aerodynamic and electroglottographic measures of normal voice production: Intrasubject variability within and across sessions. *Journal of Speech and Hearing Research, 37,* 38-45.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97,* 3099-3111.

Hombert, J. M. (1977). Consonant types, vowel height and tone in Yoruba. *Studies in African Linguistics, 8,* 173-190.

Homma, Y. (1973). An acoustic study of Japanese vowels. *Study of Sounds, 16,* 347-368.

Honda, K., & Fujimura, O. (1991). Intrinsic vowel F0 and phrase-final F0 lowering: phonological vs. biological explanations. In J. Gauffin & B. Hammarberg (Eds.), *Vocal fold physiology: acoustic, perceptual, and physiological aspects of voice mechanisms* (pp. 149-157). San Diego, CA: Singular Publishing Group.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25,* 105-113.

Iivonen, A. K. (1989). Regionally determined realization of the standard German vowel system. Mimeographed Series of the Department of Phonetics, University of Helsinki, 15, 21-28.

Ivic, P., & Lehiste, I. (1965). Prilozi ispitivanju fonetske i fonoloske prirode akcenata u savremenom srpskohravatskom jeziku: 2. *Zbornik za Filologiju i Lingvistiku, 8,* 75-117.

Kim, K. (1968). F0 variations according to consonantal environments. In Phonology Laboratory, University of California, Berkeley.

King, L., Ramming, H., Schiefer, L., & Tillmann, H. G. (1987). Initial F0-contours in Shanghai CV-syllables–an interactive function of tone, vowel height, and place and manner of stop articulation. In *Proceedings XIth International Congress of Phonetic Science, 1* (pp. 154-157). Tallinn, Estonia: Academy of Sciences of the Estonian SSR.

Kingston, J. (1993). The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech, 35*, 99-113.

Koopmans-van Beinum, F. J. (1980). *Vowel contrast reduction: an acoustic and perceptual study of Dutch vowels in various speech conditions*. Amsterdam: Academische Pers.

Ladd, D. R., & Silverman, K. E. A. (1984). Vowel intrinsic pitch in connected speech. *Phonetica, 41*, 31-40.

Ladefoged, P. (1968). *A phonetic study of West African languages: An auditory-instrumental study*. Cambridge: Cambridge University Press.

Ladefoged, P., & Maddieson, I. (1990). Vowels of the world's languages. *Journal of Phonetics, 18*, 93-122.

Lane, H., & Webster, J. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America, 89*, 859-866.

Lavoie, J. (1994). La fréquence intrinsèque des voyelles en français québécois. In F. Z. Belyazid, S. Belyazid, G. Cochrane, J. Côté, J. de Blois, M. Faucher, F. Jean, & W. Zouali, F. Z. Belyazid, S. Belyazid, G. Cochrane, J. Côté, J. de Blois, M. Faucher, F. Jean, & W. Zoualis, *Actes des huitsieme Journées de linguistique* (pp. 109-113). Quebec City: Université Laval.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America, 33*, 419-423.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 13-44). Orlando, FL: Academic.

Lounsbury, F. (1953). *Oneida verb morphology*. New Haven, CT: Yale University Press.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Meyer, E. A. (1896-7). Zur Tonbewegung des Vokals im gesprochenen und gesungenen einzelwort. *Phonetische Studien* (Beiblatt zu der Zeitschrift Die Neuren Sprachen), *10*, 1-21.

Möbius, B., Zimmermann, A., & Hess, W. (1987). Microprosodic fundamental frequency variations in German. In *Proceedings XIth International Congress of Phonetic Science, 1* (pp. 146-149). Tallinn, Estonia: Academy of Sciences of the Estonian SSR.

Mohr, B. (1971). Intrinsic variations in the speech signal. *Phonetica, 23*, 65-93.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85*, 2088-2113.

Neweklowsky, G. (1975). Spezifische Dauer und spezifische Tonhöhe der Vokale. *Phonetica, 32*, 38-60.

Nishinuma, Y. (1979). *Un modèle d'analyse automatique de la prosodie: accent et intonation en japonais*. Paris: Centre National de la Recherche Scientifique.

Nittrouer, S., McGowan, R. S., Milenkovic, P. H., & Beehler, D. (1990). Acoustic measurements of men's and wormen's voices: A study of context effects and covariation. *Journal of Speech and Hearing Research, 33*, 761-775.

Ohala, J. J., & Eukel, B. W. (1987). Explaining the intrinsic pitch of vowels. In R. Channon & L. Shockey (Eds.), *In honor of Ilse Lehiste* (pp. 207-215). Dordrecht: Foris.

Pakerys, A. (1982). *Lietuviu bendrines kalbos prozodija*. Vilnius: Mokslas.

Perkell, J., Lane, H., Svirsky, M., & Webster, J. (1992). Speech of cochlear implant patients: A longitudinal study of vowel production. *Journal of the Acoustical Society of America, 91*, 2961-2978.

Peterson, G. E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research, 4*, 10-29.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

Pettorino, M. (1987). Intrinsic pitch of vowels: an experimental study on Italian. In *Proceedings XIth International Congress of Phonetic Science, 1* (pp. 138-141). Tallinn, Estonia: Academy of Sciences of the Estonian SSR.

Pilszczikowa-Chodak, N. (1972). Tone-vowel height correlation and tone assignment in the patterns of verb and noun plurals in Hausa. *Studies in African Linguistics, 3*, 399-421.

Rakotofiringa, H. (1968). *Contributions a l'étude de la phonétique malgache II: hauteur, durée et intensité vocaliques efficaces*. Université de Grenoble.

Rakotofiringa, H. (1982). *L'accent et les unites phoniques elementaires de base en malgache-merina*. Lille: Atelier National de Reproduction des Theses, Université de Lille.

Reinholt Petersen, N. (1978). Intrinsic fundamental frequency of Danish vowels. *Journal of Phonetics, 6*, 177-189.

Rossi, M., & Autesserre, D. (1981). Movements of the hyoid and the larynx and the intrinsic frequency of vowels. *Journal of Phonetics, 9*, 233-249.

Samaras, M. (1972). Influence de l'entourage consonantique sur les variations de la frequence laryngienne des voyelles du grec moderne. *Bulletin de l'Institut de Phonétique de Grenoble, 1*, 57-66.

Sapir, S. (1989). The intrinsic pitch of vowels: Theoretical, physiological and clinical considerations. *Journal of Voice, 3*, 44-51.

Schiefer, L. (1987). F0 perturbations in Hindi. In *Proceedings XIth International Congress of Phonetic Science, 1* (pp. 150-153). Tallinn, Estonia: Academy of Sciences of the Estonian SSR.

Shadle, C. H. (1985). Intrinsic fundamental frequency of vowels in sentence context. *Journal of the Acoustical Society of America, 78*, 1562-1567.

Shi, B., & Zhang, J. (1987). Vowel intrinsic pitch in standard Chinese. In *Proceedings XIth International Congress of Phonetic Science*, 1 (pp. 142-145). Tallinn, Estonia: Academy of Sciences of the Estonian SSR.

Silverman, K. E. A. (1987). *The structure and processing of fundamental frequency contours*. Unpublished doctoral dissertation, University of Cambridge.

Sorenson, D. N. (1989). A fundamental frequency investigation of children ages 6-10 years old. *Journal of Communicative Disorders, 22*, 115-123.

Steele, S. A. (1986). Interaction of vowel F0 and prosody. *Phonetica, 43*, 92-105.

Steffen-Batóg, M. (1970). The influence of intrinsic vowel pitch on the differences in the realizations of intended intervals. *Speech Analysis and Synthesis, 2*, 177-194.

Svantesson, J. (1988). Voiceless stops and F0 in Kammu. *Lund University Department of Linguistics Working Papers, 34*, 116-119.

Svantesson, J. (1993). Phonetic correlates of register in Paraok. *Reports from Uppsala University Linguistics (RUUL), 23*, 102-105.

Tamas, S. (1976). *A beszédfolyamat alaptényezói*. Budapest: Akadémiai Kiadó.

Taylor, H. C. (1933). The fundamental pitch of English vowels. *Journal of Experimental Psychology, 16*, 565-582.

Traunmüller, H. (1982). Der vokalismus im ostmittelbairischen. *Zeitschrift fur Dialektologie und Linguistik, 3*, 289-333.

van Son, R. (1993). *Spectro-temporal features of vowel segments*. Amsterdam: IFOTT.

van Son, R. (1993). *Spectro-temporal features of vowel segments*. Amsterdam: IFOTT.

Vilkman, E., Aaltonen, O., Raimo, I., Arajärvi, P., & Oksanen, H. (1989). Articulatory hyoid-laryngeal changes vs. cricothyroid muscle activity in the control of intrinsic Fo of vowels. *Journal of Phonetics, 17*, 193-203.

Whalen, D. H., Levitt, A. G., Hsiao, P.-L., & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9- and 12-month-old French- and English-learning infants. *Journal of the Acoustical Society of America*, 97, 2533-2539.

Zawadzski, P. A., & Gilbert, H. R. (1989). Vowel fundamental frequency and articulatory position. *Journal of Phonetics, 17*, 159-166.

Zee, E. (1980). Tone and vowel quality. *Journal of Phonetics, 8*, 247-258.

Zhu, X. (1994). Shanghai tonetics. Unpublished doctoral dissertation. The Australian National University, Canberra, NSW.

FOOTNOTES

**Journal of the Acoustical Society of America*, inpress.

†Also French Department, Wellesley College, Wellesley.

[1] After this paper was submitted, we became aware of Zhu's (1994) report on Shanghai. He reports values for six men and five women as follows (taken from the end of tone 2, as we did for King et al., 1987): /u/: 131/244; /i/: 130/231; /a/: 118/217.

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 15-24

Intrinsic F0 of Vowels in the Babbling of 6-, 9- and 12-month-old French- and English-learning Infants*

D. H. Whalen, Andrea G. Levitt,† Pai-Ling Hsiao,‡ and Iris Smorodinsky†††

In every language so far examined, high vowels such as [i] and [u] tend to have higher fundamental frequencies (F0s) than low vowels such as [a]. This intrinsic F0 effect (IF0) has been found in the speech of children at various stages of development, except in the one previous study of babbling. The present study is based on a larger set of utterances from more subjects (six French- and six English-learning infants), at the ages 6, 9 and 12 months. We find, instead, that IF0 appears even in babbling. There is no indication in our data of a developmental trend for the effect, and no indication of a difference due to the target language. These results support the claim that IF0 is an automatic consequence of producing vowels.

INTRODUCTION

The relationship between vowel height and fundamental frequency (F0) has been noted for at least 60 years (Taylor, 1933). High vowels such as [i] and [u] tend to have higher F0s than low vowels such as [a] and [æ]. The mechanism for this "intrinsic F0" (IF0) or "intrinsic pitch" has been the subject of great dispute (see the reviews in Ohala & Eukel, 1987; Sapir, 1989; Fischer-Jørgensen, 1990), but the consistency of the effect is not in question (Whalen & Levitt, in press). Every language that has been examined for IF0 (31 are listed in that work) has been found to have it, and these languages represent 11 of the world's 29 major language families. IF0 has been found not only in languages such as English (Peterson & Barney, 1952) and French (DiCristo, 1982) that use F0 primarily for stress and intonation, but also in tone languages such as Mandarin (Shi & Zhang, 1987) that use F0 changes to distinguish words. IF0 seems to be insensitive to the size of the vowel inventory as well, since both small (e.g., Japanese with 5 vowels) and large (e.g., German with 14) systems show similar effects (Whalen & Levitt, 1995).

With such universality, IF0 has typically been assumed to be an automatic consequence of vowel articulation. Indeed, the theories reviewed in Sapir (1989) take this as a given. Under that assumption, it is of great interest whether the vowels of babbling will show this effect, since the babbling child presumably has no vowel categories per se, but simply vocalic articulations. If the child's vocal apparatus already has the interconnections that produce the IF0 effect in adults, then we should see IF0 in babbling. If there are significant anatomical or coordinative differences between infants and adults, perhaps IF0 will not appear in babbling.

Only one study that we have found has examined this question (Bauer, 1988). Bauer examined 3 infants at 9 and 13 months. There were 201 vowels measured at 13 months and an unreported number for the earlier age. Vowels were put into one of four broad classifications:

This research was supported by NIH grant DC-00403 to Haskins Laboratories and Catherine Best. Portions of this research were presented at the 126th Meeting of the Acoustical Society of America, Denver, Colorado, October, 1993, and the 9th International Conference on Infant Studies, Paris, June, 1994. Additional help with the stimuli was provided by Michele Sancier, Winifred McGowan, and Julia Irwin. We thank Arthur S. Abramson, Catherine T. Best, Hartmut Traunmüller, Keith Johnson, and an anonymous reviewer for helpful comments.

high front, high back, low front or low back. Bauer found no effect of height, but did find an effect of front/back. He attributed this to the high position of the larynx in the infant (Crelin, 1987). This high position also leads to a more vertical orientation, which might lead to more influence of the tongue pulling in the front/back dimension.

As a note of caution, though, the number of subjects and the number of tokens in Bauer's study were both rather small. The size of the study can greatly affect the outcome, as can be seen in a similar failure to find a vowel height effect, this time in running speech. Umeda (1981) measured approximately 200 vowels from two speakers in spontaneous conversation. She found no evidence of IF0 and concluded that it was not present in running speech. However, there are a great many factors that influence F0 in speech, and these were not controlled for in her study. To counteract this variability, it is necessary either to increase the number of observations, or to control the context. When factors such as sentence focus and segmental environment are properly controlled, even running speech shows the effect (Ladd & Silverman, 1984; Shadle, 1985). We can presume, then, that a larger sample of unrestricted text would show the effect. And, of course, it is not possible with babbling to restrict the context, so an increase in sample size is our only alternative. Thus the issue of IF0 in babbling cannot be considered to be settled, and the present study attempts to increase our understanding of this issue.

Although most researchers assume that IF0 is an automatic consequence of vowel production, others hold that IF0 is a deliberate enhancement of the speech signal by the speaker (Diehl & Kluender, 1989; Diehl, 1991). This account assumes that the perception of vowel height is a function not only of F1 frequency but of the difference between F0 and F1 (Traunmüller, 1981) and that speakers intentionally increase their F0 with high vowels to make this difference larger than it would otherwise have been. The universality of IF0, on this account, only argues for the usefulness of this particular enhancement. In babbling, however, there is no communicative intent and thus no distinctions to enhance. So the enhancement account should predict that IF0 will not appear in babbling. Even Bauer's (1988) finding, if it is correct, would be inconsistent with the enhancement account, since it implies an automatic (though different) mechanism for IF0. If IF0 were not found for babbling, the enhancement account would seem to be supported, with the assumption that IF0 would be an enhancement acquired later in development.

If IF0 is found for babbling, the most likely explanation is that it is not only universal but automatic. For the enhancement account to accommodate such a result, it would seem that an imitative explanation would be necessary. That is, since children hear this vowel height/F0 correlation in whatever adult language they hear, they include it in their babbling. Enhancement per se should not be an issue, since there are (presumably) no categories to enhance, but the imitation might be complex enough to include small F0 changes. This issue will be addressed further in the Discussion.

If enhancement is operative, we might expect there to be a developmental trend toward increased usage of the enhancement. Previous studies of IF0 in older children, with ages ranging from 5 to 11 years, show no indication of a developmental trend in IF0. Table 1 presents results from five published studies (Peterson & Barney, 1952; Peterson, 1961; Sorenson, 1989; Glaze, Bless, & Susser, 1990; Hillenbrand, Getty, Clark, & Wheeler, in press). We have averaged the two high vowels [i] and [u] and the two low vowels [a] and [æ]. As can be seen in the difference column of Table 1, there is variability, especially in the Sorenson values where the N was small (only 3 per cell). But there is no indication of an overall trend toward larger (or smaller) effects.

If IF0 is universal, then we would expect to find similar patterns in the babbling of infants from any language environment. If IF0 is deliberate enhancement, we might expect that different languages would use the enhancement to different degrees. This difference might then appear as a difference in the babbling behavior of children in different language communities. The present study takes a first step in assessing the universality of IF0 in babbling by examining infants in two language environments, English and French. We have already found intonational differences between these two language groups in an earlier study (Whalen, Levitt, & Wang, 1991). That study included 10 of the 12 subjects analyzed here. Since these children are using F0 in different ways in their babbling, it is certainly possible that they would treat IF0 differently if they were producing IF0 deliberately.

IF0 in babbling, then, needs further examination. The present study examines the babbling of 12 infants, 6 each in English and French environments. The infants were recorded in the home at 6, 9 and 12 months of age.

Table 1. *Average F0 values for high versus low vowels for five studies that include children. Adult values for English (Peterson and Barney, 1952) and French (DiCristo, 1982) are given for comparison. Age is in years. F0 is in Hz. The difference is given both as a Hz value and (in parentheses) as a percentage of the a/æ value.*

Study	Age (yrs)	N	Sex	F0 for i/u	F0 for a/æ	Difference
DiCristo, 1982	adult	1	fem	239	226	13 (5.8)
	adult	3	male	133	124	9 (7.3)
Peterson and Barney, 1952	adult	33	male	139	126	13 (10.3)
	adult	28	fem	233	211	22 (10.4)
	"child"	15	both	274	253	21 (8.3)
Peterson, 1961	"child"	3	both	294	262	32 (12.2)
Glaze et al., 1990	5-11	97	both	250	229	21 (9.2)
Hillenbrand et al., in press	10-12	46	both	248	229	19 (8.3)
Sorenson, 1989	6	3	male	290	258	32 (12.4)
	6	3	fem	324	301	23 (7.6)
	7	3	male	307	288	19 (6.6)
	7	3	fem	288	279	9 (3.2)
	8	3	male	267	255	12 (4.7)
	8	3	fem	286	264	22 (8.3)
	9	3	male	272	229	43 (18.8)
	9	3	fem	300	275	25 (9.1)
	10	3	male	263	243	20 (8.2)
	10	3	fem	273	279	-6 (-2.1)
(average)	6-10	30	both	287	267	20 (7.5)

We measured all the vowels except for the central (e.g., [ə]) and lower-mid (e.g., [ɛ] and [ɔ]). This resulted in 7,325 tokens to analyze. With a larger set of results, we can more confidently address the issue of whether IF0 is automatic or under the speaker's control.

II. Methods

A. Subjects. The speakers were 12 infants, 6 learning French as their native language and 6 learning American English. The French infants were all living in Paris or its environs. The American infants lived in various cities on the northeast coast of the United States.

B. Stimuli. The utterances for the present study were selected from recordings made at weekly intervals by the parents of the children. Each infant was recorded in the home on a cassette tape recorder (Panasonic RQ 3145 or Marantz PMD 430) using a high quality microphone (Realisitc supercardioid 33992A). Individual recording sessions lasted approximately 10-20 minutes. The parents were asked to choose a time when the child was likely to be alert and unlikely to cry. As far as possible, the microphone was held 20 cm from the baby. If necessary, the parent could attempt to induce babbling by speaking to the child (stopping, of course, when the infant began vocalizing). Additional comments about the session were recorded by the parent on a form provided with each tape.

All utterances from the 6, 9 and 12 month tapes were digitized onto the Haskins Laboratories VAX computer system. They were low-pass filtered at 9.6 kHz and sampled at 20 kHz, with pre-emphasis (Whalen, Wiley, Rubin, & Cooper, 1990). We excluded cries, whispers, and various vegetative sounds. If an utterance contained a combination of speech and nonspeech, we would try to transcribe the speech.

All the utterances were transcribed by the third author, a native speaker of Mandarin Chinese. He is phonetically trained, and has experience with a wide variety of languages. Transcriptions were made from the digitized waveform, with the help either of the Haskins Laboratories program HADES (Rubin, 1995) or Signalyze® (Keller, 1990). With these programs, the whole utterance could be heard repeatedly, as could any selected portion of the utterance. The overall character of the waveform also gave indications of possible syllables. The symbols of the International Phonetic Alphabet were used, with the understanding that some of the utterances would be very difficult to transcribe. We felt that obtaining a more detailed transcription was worth the effort

BEST COPY AVAILABLE

involved, since this allows us to make more comparisons than Bauer's (1988) four-way classification.

Once the transcriptions were made, we selected the following vowels for analysis: high front ([i y ɪ ʏ]), mid front ([e ø]), low front ([æ a ɑ]), high back ([u ɯ ʊ]), mid back ([o ɤ]) and low back ([ɑ ɒ ɐ]). (Strictly speaking, [ɐ] is low central, but there were few enough members of this group anyway, so we included it.) While we did want to have the most accurate transcription possible, it was not possible to analyze the results any more finely than this, primarily because of the small number of instances of many of the vowels. There were a handful of tokens that were nasalized; these were simply included without any indication of the nasalization. We also treated all vowels without regard to their consonantal environment.

C. Analysis. All fundamental frequencies were measured from the speech waveform, by hand, using either HADES or Signalyze®. The following procedure was used: For each syllable containing one of the vowels of interest, the main period of vowel activity was delimited. Then, a location 40% of the way into this segment was found. In the best case, we would then measure five pitch periods to the left of the point and five to the right. The duration of this ten pitch period segment was then translated into an average F0 for that measurement point. In some cases, the pitch periods immediately around the 40% point were not measurable, either because the waveform was noisy or low in amplitude, or otherwise unclear. In those cases, the nearest 10 measurable pitch periods within that syllable were chosen. Some tokens that had been transcribed proved to be too noisy or too faint to measure. Table 2 presents the number of measured tokens for the 12 subjects at the three ages. Two of the subjects lacked recordings at some of the months: JZ was missing the 9 and 12 month recordings, and MB was missing the 12 month recording. Both were French subjects. Another French subject, YC, lacked 12 month recordings but had 11 month ones. The 11 month recordings were used for the 12 month data for her. Two other subjects had sparse data at one or two months, so these were supplemented with recordings from an adjacent month. For English subject MA, 25.8% of the 6 month data was from the 6 month recordings, while the remaining 74.2% came from the 7 month recordings. Also for this subject, 37.7% of the 12 month data was from the 12 month recording, while the remaining 62.3% was from the 11 month recordings. Finally, for French subject MB, 35.0% of the 6 month data came from the 6 month recordings, while the remaining 65.0% came from the 7 month recordings.

All of the target vowels were measured, with an exception for one subject. American subject NG had a large corpus, but the vast majority of her vowels were [e]. At six months, [e] was approximately 45 times as frequent as the next vowel. At nine months, the ratio was around 17 to 1. By 12 months, [e] outnumbered its nearest rival by a mere factor of 10. In order both to keep the representations of the vowels relatively similar, and to cut down on the amount of work required for this subject, only selected [e]'s were analyzed for her. For each age, a number of [e]'s was counted out (45, 17 or 10, for the three ages). The utterance containing that [e] was analyzed for all its [e]'s. Thus if there was only one [e], then that would be the only one analyzed. If that utterance happened to have several [e]'s, all of them were analyzed. In this way, [e] was still the most frequent vowel, but only by an overall factor of 2.3.

The distribution of these vowels is similar to those found in previous studies. In one cross language study (Boysson-Bardies, Hallé, Sagart, & Durand, 1989), back vowels were found to be relatively rare (6.6% of the utterances in French and English), though the low back vowels were the most common of those. In the present study, by contrast, the high back vowels accounted for a higher proportion of the back vowels than was the case for the other study. (The proportion of front vowels overall would be higher if we had not excluded many of NG's [e] vowels.)

Table 2. *Number of tokens analyzed for the 12 subjects. Speaker JZ had no recordings at 9 and 12 months, and MB had none at 12 months.*

Language	French						English					
Initials	MS	NM	YC	JZ	EC	MB	MM	VB	MA	AB	CR	NG
Sex	M	F	F	M	M	F	M	F	M	F	M	F
# at 6 mo.	197	10	161	278	72	95	28	293	61	1250	80	130
# at 9 mo.	105	33	353	–	37	83	38	123	30	708	130	337
# at 12 mo.	147	29	850	–	171	–	48	205	226	647	136	234

Our proportions are more in agreement with Buhr's (1980) one English-learning infant. The selection criteria used here were too different to allow a direct comparison with the de Boysson-Bardies et al. (1989) study, but the distribution of vowels analyzed here is at least qualitatively similar to that found in other studies.

F0s larger than 700 Hz were excluded from the analysis. These represented 4.3% of the 7651 tokens measured. Such extreme values, while common in babbling, distort the means for those cells with small Ns. It is also possible that a different phonation type is involved in such high F0s, which would be a second reason to exclude these values. The selection process resulted in 7,325 tokens being measured for the 12 subjects.

II. Results

Means for the six vowel types for the twelve subjects are given in Table 3. Also given are the number of tokens that went into each value. Table 4 gives the size of the IF0 effect, both for height and front/back. The front/back difference is given as the front vowel mean minus the back vowel mean. In this way, any difference that matches the results of Bauer (1988) will be positive in value, while contrary results will be negative. As can be seen, there is a positive difference for height for 10 of the 12 subjects. For front/back, only five of the 12 subjects match Bauer's results.

For an analysis of variance, we operated on the means for each of the six cells for the 12 subjects. An analysis that used each observation was attempted, but the enormously large degree of freedom for the error term meant that almost any difference, however trivial, appeared significant. Using the means also gives the subjects with fewer productions a stronger say in the analysis. Since the differences among speakers, not tokens, are of primary importance, this result is to be desired.

The analysis, then, included the grouping variable Language (English or French), and two within factors, Height and Front/Back (with 3 and 2 levels, respectively). Three of the subjects (MM, NM, and MB) have missing cells, due to the lack of any instances of the mid back vowel category. Rather than reject these subjects from the analysis, these cells were replaced with the means of the five other cells for these subjects. This is a conservative approach to data replacement, since it will tend to minimize differences that actually exist. Language was not a significant main effect ($F(1,10) < 1$, n.s.), indicating that the babblers had roughly equivalent overall F0s. Height was a significant factor ($F(2,20) = 16.62$, $p < .001$), while the interaction with language was not ($F(2,20) = 2.09$, n.s.). Front/Back was also not a significant factor ($F(1,10) < 1$, n.s.); neither was the interaction with language ($F(1,10) < 1$, n.s.). The two-way interaction of Height and Front/Back was significant ($F(2,20) = 4.48$, $p < .05$), but the three-way interaction with Language was not ($F(1,10) = 1.09$, n.s.).

Table 3. *F0 values for the six vowel categories for the 12 subjects.*

	Front			Back		
	Mean F0	N	% of total	Mean F0	N	% of total
High	405.7	519	7.1	381.8	302	4.1
Mid	364.6	5254	71.7	359.9	89	1.2
Low	332.2	808	11.0	330.6	353	4.8

Table 4. *Size of the high/low difference and the front/back difference in F0 for the 12 subjects. The first and third rows are in Hz, the middle row is the high/low difference expressed as a percentage of the low vowel F0.*

	French						English					
Initials	MS	NM	YC	JZ	EC	MB	MM	VB	MA	AB	CR	NG
High - Low	56.0	94.3	40.4	18.4	-9.3	-26.7	15.6	70.9	15.1	93.8	101.3	81.4
(as %)	16.8	23.9	12.6	5.1	-2.7	-6.4	3.7	21.0	4.2	27.0	35.8	26.0
Front - Back	-11.1	-19.5	-16.7	-13.9	19.3	-35.7	20.1	-28.7	-15.4	34.3	7.0	54.9

For the analysis by age, it was necessary to restrict the number of cells. By the time we break the results down into the six categories and the three ages for the twelve subjects, 45 of the 216 cells are empty. Most of these are for the low back and mid back vowels. Therefore, we analyzed the four other cells, as a single factor of Vowel Quality with four levels, so that only differences among the four cells can be tested, not the front/back and high/low dimensions. Eighteen of the 45 empty cells come from the two subjects who lacked certain months, as mentioned before: JZ had no 9 or 12 month data, and MB had no 12 month data. These two subjects were excluded from this analysis, so that the remaining subjects had no missing cells. The ANOVA factors were Vowel, with four levels, Language, with two levels (English and French), and Age, with three levels (6, 9 and 12 months).

In this analysis, as before, Language was not a significant factor ($F(1,8) < 1$, n.s.). It did not enter into any significant interactions either. Age was not a significant main effect ($F(2,16) < 1$, n.s.), which is to be expected: Even though F0 lowers throughout development (see Table 1), the time elapsed here is too short to show this effect. Vowel is a significant main effect ($F(3,24) = 8.22$, $p < .001$), again showing the height effect. The critical interaction, Age by Vowel, is not significant ($F(6,48) = 1.97$, n.s.), giving no indication of a difference in the effect over the six months involved here (see Table 5). Even if we analyze each month separately (despite the lack of an interaction), the IF0 effect is present at each age. The separate analyses are strong for the 6 month ($F(3,27) = 9.61$, $p < .001$) and 12 month ($F(3,27) = 5.05$, $p < .01$) measurements, and somewhat less robust for the 9 month ($F(3,27) = 2.94$, $p = .0512$). There is no evidence of change in IF0 over this time span.

Since we relied on our transcriptions to separate the vowels into categories, we need to be sure that we can do this independently of F0. In adult speech, it is certainly clear that different vowels can be produced with a wide range of F0s without losing the vowel's identity. With babbling, however, it is not possible to ask the speaker to reproduce a particular vowel. One way of avoiding the vowel identity problem would be to correlate F0 with F1. Since F1 is lower with the high vowels and higher with the low vowels, there should be a negative correlation between F0 and F1 when IF0 is present. In the Peterson and Barney (1952) data, in fact, there is such a correlation if we examine the three speaker groups (the 33 adult males, the 28 adult females, the 15 children) separately. When we correlate each individual production (there were two per vowel) for each vowel for all the speakers, we obtain the following correlations: males, $r = -.16$ ($p < .001$), females, $r = -.20$ ($p < .001$), children, $r = -.10$ ($p < .10$). The correlation does not reach significance for the children either because of greater variability of their values or the smaller number of subjects.

When we examined our babbling data, however, there was a positive correlation between F0 and F1, but this was due to the fact that the formants were almost invariably excited by a single harmonic. With a mean F0 of 370 Hz, the formants in our set of babbles are poorly represented. If a harmonic happens to be at the center frequency of a formant, the two nearest harmonics would be approximately 15 db lower in amplitude even with a bandwidth of 100 Hz. (For adults, bandwidths typically remain in the range 50-60 Hz for formant values up to 2000 Hz (Dunn, 1961).) Harmonics with such low amplitudes are too close to the background level to contribute to the measurement of the formant. Occasionally in our measurements, we found two harmonics of equal amplitude, and it was possible to assume that the center frequency of the formant was between them.

Table 5. *Mean F0s for four of the six vowel categories for the 10 subjects that had measurements at each of the months analyzed. The mid-back and low-back categories were missing for many of the subjects for one month or another. The last row shows the difference between the mean of the two high vowel categories and the low vowels category.*

Age in Months	6		9		12	
	Mean F0	N	Mean F0	N	Mean F0	N
High Back	373.5	88	369.8	71	394.6	125
High Front	402.3	115	412.5	171	403.0	221
Mid Front	350.4	1757	381.8	1311	364.1	1838
Low Front	313.6	227	336.7	190	336.9	327
Diff. between Highs and Low	74.3		54.5		61.9	

(With such limited measurements of the formant frequencies, it was, of course, impossible to measure the bandwidth with any confidence.) The appearance of two harmonics was uncommon, so the formant value was much more likely to be identical to one of the harmonics, resulting in the positive correlation between F0 and formant frequency.

As a final check on the possible misperception of F0 as vowel quality, we examined the distribution of the vowel categories by F0. If F0 were the only factor, then the distributions should be distinct. If any vowel can occur on any F0, then the distributions should be greatly overlapped. Figure 1 shows a highly overlapped pattern. For that figure, the number of tokens of a particular vowel in an F0 range or "bin" of approximately 16 Hz was counted. The top panel shows the proportion of all the vowels in the six categories. Because the mid front category is so disproportionately represented, it is hard to see the other distributions. Thus the lower panel shows the same data with a ceiling on the mid fronts. As is clear, there are vowels of each category at every level of F0. Certainly the distributions are different, since that is what the IF0 effect consists of. But it is not the case that a high F0 was enough to cause a perception of a high vowel. The identifiability of the vowels was apparent throughout the F0 range. Thus there is no evidence of any large perceptual bias in the transcriptions.

It is impossible to rule out smaller perceptual biases which might have influenced the results. Indeed, small effects of F0 on the identification of ambiguous vowels have been found in one study by Reinholt Petersen (1986). Using synthetic vowels ranging from [u] to [o], he found that the most ambiguous vowel received more [u] responses with a high F0 compared with the low F0. The effects were quite small and never enough to change the majority decision. In addition, it was only possible to shift an ambiguous vowel from one category to a neighboring category. The results of Gottfried and Chew (1986), in which a wide range of F0s for sung vowels was used, also show extremely few instances in which the height of the perceived vowel differs by more than one level. Thus even if there were bias effects in the present transcriptions, such biases would not account for the F0 difference between the low vowels and the high vowels. Despite the impossibility of completely ruling out small bias effects, then, the pattern of results strongly suggests that the effects we have found are due to the vowel articulation and not to the transcription.

III. Discussion

Our analysis of the babbling of 12 infants, six each from English- and French-learning situations, indicates that the intrinsic F0 (IF0) associated with vowels appears even in babbling. There was no evidence of a change in the effect across the three ages examined (6, 9 and 12 months). There was also no evidence of any difference between the two languages. These results are most compatible with the hypothesis that IF0 is an automatic consequence of vowel production.

The previous study that examined this question (Bauer, 1988) did not find a height difference, but instead found a front/back difference. That author attributed the fact that infants differ from adults to the relatively high position of the larynx in the vocal tract of young children (Crelin, 1987). However, we believe that his results differ from ours because of the scope of the studies. Bauer examined three children, and only 201 tokens at age 13 months. (He does not report the number of tokens for the 9-12 month portion.) This is too small a number to use for an unconstrained situation such as babbling. If we could have infants give us multiple repetitions with the same intonation, then a smaller number would be enough. But infants are constantly exploring the F0 range as they babble, and the placement of vowels of different qualities is random in this distribution. Thus it is very easy to have several utterances with high, even squealy pitch with a low vowel. It takes a large sample for this to average out. As can be seen in Table 4, there were two subjects who showed higher F0s for low vowels, and they were two among those with the smallest number of tokens to analyze.

The differences between front and back vowels were not consistent from subject to subject in the present study. This is unlike Bauer's (1988) results but like the adult studies (Whalen & Levitt, in press). Given that Bauer's explanation of the front/back effect as due to the high larynx position is a plausible one, we need to explain why this high position does not change the IF0 effect. In fact, the high position seen in Crelin's x-ray images is somewhat misleading, since most of those were taken at rest. As Crelin himself notes (1987:96), the larynx is pulled down into a much more adult-like position during speech (and screaming). That is, infants must work to make their vocal tracts appear more adult-like, and

doing so seems to bring their larynx into the same relationship with the tongue that adults have. Since they show the same IF0 as adults, it seems likely that the same mechanism is involved as well. One might still suppose that different children might adopt different strategies, but even the two subjects who showed a contrary effect for height were inconsistent for front/back: Subject EC had the difference that Bauer found, while MB went in the other direction. So, as with the adults, there is no consistent effect of the front/back dimension on F0.

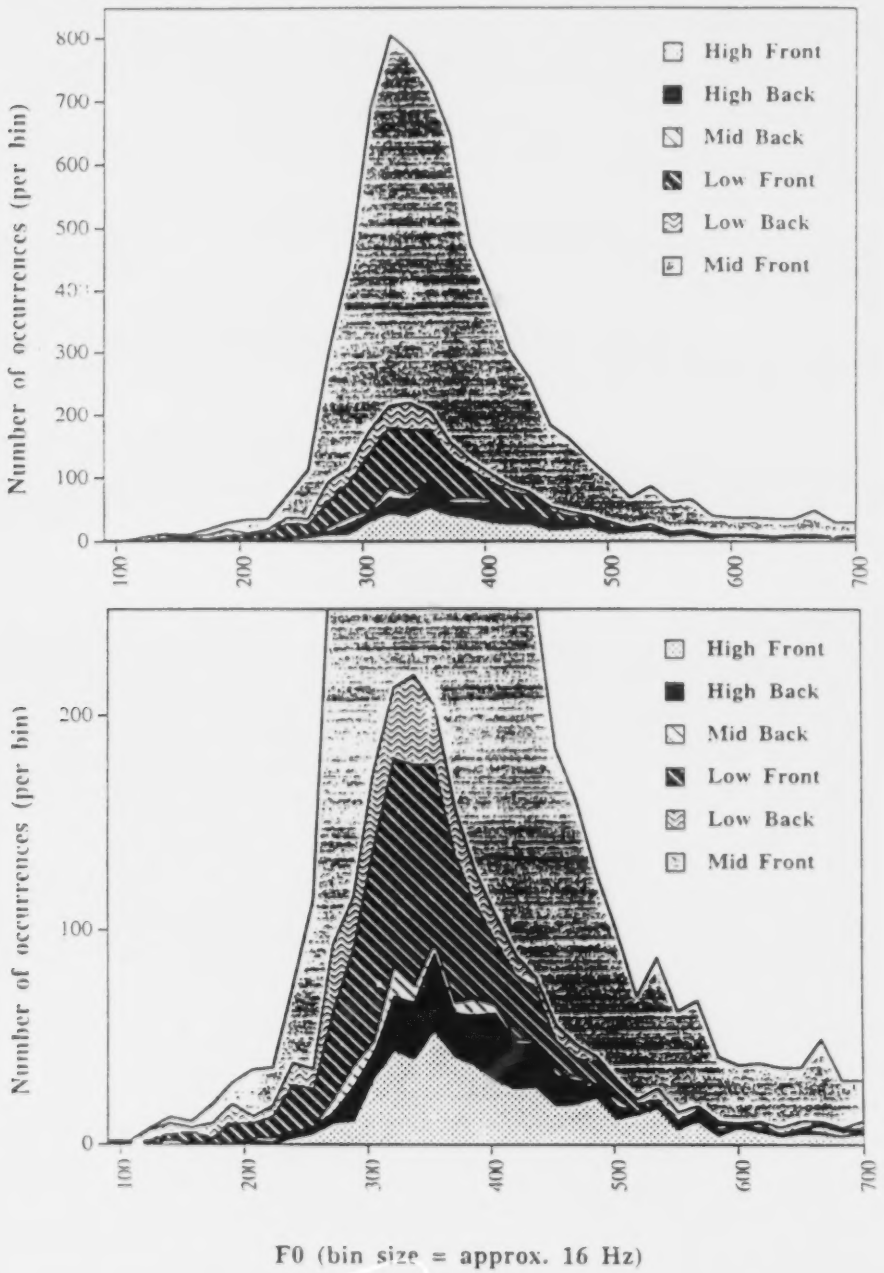


Figure 1. **Distribution of the six vowel categories by F0. Top panel: all vowels analyzed. Lower panel: the same data truncated at 250 occurrences, giving better resolution for the less well represented categories.**

The present study also found no evidence of a developmental change over the six month span examined. This is consistent with the universality of IF0 (Whalen and Levitt, in press) and with the lack of any evidence of a developmental change later in life (Table 1). The overall percentage of difference found for the babblers was 13.9% (as calculated from Table 3). This is slightly higher than that found for other studies (Table 1), but a statistical artifact is probably the cause. If we had transformed the measured F0s into a semitone scale before averaging, the effect of the very high F0s would have been reduced, and the difference between high and low vowels would probably have been much more similar, too. Certainly, if there is any developmental trend, it is for less IF0 rather than more. This does not fit with the enhancement hypothesis.

The present results are at odds with the one previous study (Bauer, 1988) and call into question the explanation given there. The difference is most likely due to the difference in sample size (200+ tokens versus the present 7,000+). In addition, his developmental change was based on a comparison of the size of two different F ratios, which is not a reliable method of comparing results across data sets. It is also risky to assume that the absence of a significant difference means that there is no effect. However, the measurements here are sufficiently strong to show the IF0 effect at each of the three ages analyzed, so at least we know that the effect is not absent at any of the ages. It seems likeliest that there is no developmental change in IF0.

These results also cast doubt on the description of IF0 as a deliberate enhancement of the speech signal (Diehl & Kluender, 1989; Diehl, 1991), for three reasons. First, while infants may begin to perceive the vowel categories of their target language at an early age (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), the evidence from vowel productions in babbling shows only a tendency toward the target language formant space (Boysson-Bardies et al., 1989) or intonation (Whalen et al., 1991), not toward specific categories If the infants have no categories to enhance, why should they use IF0? It is true that infants will hear vowels with IF0, since every language shows IF0 (Whalen and Levitt, in press). So if they are imitating what they hear, then they might imitate IF0 differences. Infants do not imitate just the native vowel categories, though. They produce some nonnative vowels and seldom if ever produce some of the native vowels. It is hard to see why they should imitate the IF0 feature, nor indeed how they might abstract away from the individual vowels to the more general principle that vowel height is what is important. Second, the IF0 effect disappears at the lower portion of adult speakers' ranges (Whalen & Levitt, in press). The explanation for this phenomenon is likely to come from the different mechanisms for raising and lowering F0. Unless the infants have already understood this difference, the lack of IF0 at low values would seem to add uncertainty to the generalization that imitative IF0 would have to be based on. Finally, the task of detecting IF0 in the course of running speech would seem especially difficult in the case of learners of a tone language, since they certainly hear a great deal of F0 variation that is important (the tones) that is completely unrelated to IF0. However, IF0 also occurs in tone languages like Mandarin (Shi & Zhang, 1987). We might expect that infants who are learning Mandarin as their native language would fail to use IF0 if any babblers would. Given the importance of tone in Mandarin and its independence from IF0, Mandarin-learning infants might not easily produce IF0 if it had to be learned. On the other hand, if IF0 is automatic, then even the Mandarin-learning infants would show IF0. This remains to be tested.

The present study shows IF0 in babbling. The effect is independent of which of the two target languages (French or English) were involved. There does not seem to be a developmental trend for the 6 to 12 month range examined. So, despite the fact that vowels can be produced with a wide range of F0s, it appears that IF0 is an automatic consequence of vowel articulation.

REFERENCES

Bauer, H. R. (1988). Vowel intrinsic pitch in infants. *Folia phoniatrica, 40,* 138-146.

Boysson-Bardies, B. d., Halle, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling, *Journal of Child Language, 16,* 1-17.

Buhr, R. D. (1980). The emergence of vowels in an infant, *Journal of Speech and Hearing Research, 23,* 73-94.

Crelin, E. S. (1987). *The human vocal tract: anatomy, function, development, and evolution.* New York: Vantage.

DiCristo, A. (1982). *Prolégomènes a l'étude de l'intonation. micromélodie.* (Editions du Centre National de La Recherche Scientifique, Paris) .

Diehl, R. L. (1991). The role of phonetics with the study of language. *Phonetica, 48,* 120-134.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1,* 121-144.

Dunn, H. K. (1961). Methods of measuring vowel formant bandwidths. *Journal of the Acoustical Society of America, 33,* 1737-1746.

Fischer-Jørgensen, E. (1990). Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica, 47,* 99-140.

Glaze, L. E., Bless, D. M., & Susser, R. D. (1990). Acoustic analysis of vowel and loudness differences in children's voice. *Journal of Voice, 4,* 37-44.

Gottfried, T. L., & Chew, S. L. (1986). Intelligibility of vowels sung by a countertenor. *Journal of the Acoustical Society of America, 79,* 124-130.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (in press). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America.*

Keller, E. (1990). *Signalyze, signal analysis for speech and music: User's manual.* Rosemere, Quebec: InfoSignal, Inc.

Kuhl, P. J., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606-608.

Ladd, D. R., & Silverman, K. E. A. (1984). Vowel intrinsic pitch in connected speech. *Phonetica, 41,* 31-40.

Ohala, J. J., & Eukel, B. W. (1987). Explaining the intrinsic pitch of vowels. In R. Channon & L. Shockey (Eds.), *In honor of Ilse Lehiste* (pp. 207-215). Foris: Dordrecht.

Peterson, G. E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research, 4,* 10-29.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels, *Journal of the Acoustical Society of America, 24,* 175-184.

Reinholt Petersen, N. (1986). Perceptual compensation for segmentally conditioned fundamental frequency perturbation, *Phonetica, 43,* 31-42.

Rubin, P. E. (1995). HADES: A case study of the development of a signal analysis system. In R. Bennett, S. L. Greenspan, & A. Syrdal (Eds.), *Behavioral aspects of speech technology: Theory and applications* (pp. 501-520). Amsterdam: Elsevier.

Sapir, S. (1989). The intrinsic pitch of vowels: Theoretical, physiological and clinical considerations. *Journal of Voice, 3,* 44-51.

Shadle, C. H. (1985). Intrinsic fundamental frequency of vowels in sentence context. *Journal of the Acoustical Society of America, 78,* 1562-1567.

Shi, B., & Zhang, J. (1987). Vowel intrinsic pitch in standard Chinese. In *Proceedings XIth International Congress of Phonetic Science,* 1 (pp. 142-145). Tallinn, Estonia: Academy of Sciences of the Estonian SSR.

Sorenson, D. N. (1989). A fundamental frequency investigation of children ages 6-10 years old. *Journal of Communication Disorders, 22,* 115-123.

Taylor, H. C. (1933). The fundamental pitch of English vowels. *Journal of Experimental Psychology, 16,* 565-582.

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of Acoustical Society of America, 69,* 1465-1475.

Umeda, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech, *Journal of the Acoustical Society of America, 70,* 350-355.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics, 23,* 349-366.

Whalen, D. H., Levitt, A. G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language, 18,* 501-516.

Whalen, D. H., Wiley, E. R., Rubin, P. E., & Cooper, F. S. (1990). The Haskins Laboratories' pulse code modulation (PCM) system. *Behavior Research Methods, Instruments, and Computers, 22,* 550-559.

FOOTNOTES

Journal of the Acoustical Society of America, 97, 2533-2539 (1995).

†Also Department of French, Wellesley College.

‡Also Department of Linguistics, University of Connecticut, Storrs.

†††Also Department of Linguistics, Yale University.

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 25-29

Knowledge from Speech Production Used in Speech Technology: Articulatory Synthesis*

Richard S. McGowan

INTRODUCTION

There appears to be a continuing trend toward incorporating knowledge of speech production into speech technology—text-to-speech synthesis (e.g., Bickley, Stevens, & Williams, 1994; Parthasarthy & Coker, 1992), low bit rate coding (see Schroeter & Sondhi, 1992), and automatic speech recognition (e.g., Rose, Schroeter, & Sondhi, 1994; Shirai & Kobayashi, 1986). For automatic speech recognition, using knowledge of the coordination of the vocal tract articulators and the resulting acoustics can reduce apparent token-to-token variability so that general pattern recognition algorithms have less work to do. Using articulatory representations in speech coding has the potential of greatly reducing bit rate because the articulators move relatively slowly and may be described by a few parameters by using an underlying dynamical model or by using simple curve fitting. Finally, text-to-speech synthesis can be improved using articulator control parameters, because the laws of physics can be used to produce the correct bundle of acoustic features with a comparatively limited parameterization—the acoustic output is constrained by the laws of physics. All these applications that depend on articulatory representation of speech production, can be grounded in what is called an articulatory synthesizer. An articulatory synthesizer is a device that produces speech output from a set of articulatory parameters (an articulatory representation). These devices are usually implemented in software on a digital computer.

This work was supported by grant NIH grant DC-01247 to Haskins Laboratories. The description of task dynamics has benefited by discussions the author has had with Elliot Saltzman. Thanks to Phil Rubin and Doug Whalen for reviewing this work.

The production of speech using an articulatory synthesizer (the "forward" mapping) can be divided into two major components: that of finding the mapping from the linguistic units to the articulatory movement, and that of finding the mapping from the articulatory movement to the aerodynamic state and acoustic output. The forward problem is solved with the composite mapping. The first mapping is the domain of people interested in the control of human movement and coordination as it relates to the vocal tract during speech, which includes some linguists and some experimental psychologists. The second mapping from articulatory movement to aerodynamic state and acoustics is the domain of acousticians. It is easily seen that both components of the forward mapping are important for the named technical applications. To perform text-to-speech synthesis from an articulatory point-of-view, the composite mapping is constructed, and to perform low bit rate coding or automatic speech recognition, one would need to find the inverse of the composite mapping, if the approach is to use articulatory information. If an analysis-by-synthesis procedure is used to construct the inverse composite mapping, then it is necessary to construct each component forward mapping.

TASK DYNAMICS

Only one example of one part of mapping from linguistic units to articulatory movement will be discussed here. This example is the model of articulatory coordination for articulators in performing speech gestures used at Haskins Laboratories, known as task dynamics (Saltzman & Munhall, 1989). This model describes the formation and breaking of constrictions in the vocal tract using a set of independent, linear, second-order differential equations: one equation

for each constriction (e.g., labial, tongue body, or tongue tip). Because constrictions can be made with the coordinated activity of vocal tract articulators, such as lips, jaw and tongue, these equations are transformed into a model for articulator geometry. In the articulator coordinate system, the equations for constriction dynamics become coupled and nonlinear, and the pseudo-inverse of the Jacobian is used in their solution, because there are more articulator than constriction degrees-of-freedom. This means that several articulators can be used to attain the same constriction target (upper lip, lower lip, or jaw can be used to attain lip closure).

Task dynamics models phenomena that are observed in real speech behavior. This includes *articulatory compensation*, where one articulator compensates for another that cannot move. (e.g., the lips can increase their total movement to close the mouth when the jaw cannot move.) The other pervasive phenomenon in speech that task dynamics models is that of *coarticulation*. For instance the jaw can be used to close the mouth or it can be used to lower the body of the tongue for certain vowels, such as /a/. When there is mouth closure, say for /b/, followed by an /a/ both goals influence the jaw, so that the mouth closure is probably attained by more lip involvement than would be without the presence of the /a/. This is so the jaw can be lower for the following /a/.

There are some real advantages to using task dynamics in the technical applications to be considered. The first is that constrictions of the vocal tract and the output acoustics are closely related so that the analysis-by-synthesis that recovers task dynamic parameters from speech is facilitated. Further, phonology based on articulatory gestures and instantiated in task dynamics is being constructed by linguists (Browman & Goldstein, 1990). This kind of work is necessary to map the task-dynamic parameters, such as the natural frequency of a lip closure, finally to linguistic units. This, of course, is required if automatic speech recognition is to be done using an articulatory representation.

Task dynamics takes the approach of finding the appropriate coordinate system to define speech behaviors (currently, constriction dynamics) and a means of transforming this coordinate system into a physical coordinate system (vocal tract articulators). This approach is extremely valuable in attempting the man-machine applications that are named above. However, there is room for an evolution in the details of this approach. For instance, the aerodynamics of the vocal tract appear to be controlled in a task specific way, and thus these must be included in some way (McGowan & Saltzman, in press). It is not clear whether all vocal tract gestures use constriction targets, and, in particular, vowels may need a more spatially global specification (Mattingly, 1990). Also, even where constriction targets are appropriate, there may be a region of targets rather than a point target (Guenther, 1994). The extension of this model should be undertaken to account for a variety of individual vocal tract shapes and for the sequencing of gestures, which is important for the rhythm mechanisms for rate and stress.

ARTICULATION-TO-ACOUSTICS

Where are we now in terms of the mapping from articulation to acoustics in articulatory synthesis, which is the second mapping that has to be constructed? What is the relation between the physics of fluid flow in the vocal tract and the propagation models that we are currently using? All the articulatory synthesizers known to the author use one-dimensional models of wave propagation (some with corrections for large area changes). The voice source is generated in a variety of ways, including simulations of self-oscillating vocal folds. The noise sources in the vocal tract are modeled as point sources, and their amplitude and frequency characteristics depend on aerodynamics in various degrees of sophistication.

Some synthesizers are time-domain synthesizers (e.g., Maeda, 1982), so that the waves created by the sources are propagated on a space-time grid. Other synthesizers use a frequency-domain transfer function to represent the wave propagation in the vocal tract (e.g., Sondhi & Schroeter, 1987; Davies, McGowan, & Shadle, 1993). The output speech can be calculated by mapping the transfer function to the corresponding time-domain transfer function via an inverse discrete Fourier transform (DFT). An alternative is to find the poles and zeros of the transfer function and to use a formant synthesizer to produce the output speech (e.g., Lin, 1994). The remainder of the paper will suggest two research directions for articulation-to-acoustics mapping. The first is a proposal to use a set of orthonormal bases functions, other than circular functions, to represent the vocal tract transfer function. These bases functions are from what Coifman (1991, p. 881) calls a "library of wavelet packets", including wavelets, used in multiresolution analyses. The other proposal is to provide a four-parameter, articulatory model for the control of the voice source.

MULTIRESOLUTION SYNTHESIS

When a frequency domain transfer function is used to represent vocal tract wave propagation, a time domain transfer function is calculated as an inverse discrete Fourier transform (DFT), and the sources convolved with the resulting transfer function. To obtain reasonable frequency resolution, it is necessary that the transform window be of reasonable duration (25.6 ms for Sondhi & Schroeter [1987] for a 20kHz sampling rate). However, in performing the inverse transform, the vocal tract is assumed to be unchanging within the duration of the transform window; thus invoking the quasisteady (stationarity) approximation. There are speech environments for which this approximation may be inappropriate, including the closure and release of stops, fricatives, affricates, and approximants. Specifically, there are two possible problems in these speech environments. First, the filtering properties of the vocal tract may be rapidly varying, and, second, the source properties may be changing rapidly because of vocal tract changes. The voice source is affected by the configuration of the upper vocal tract in what is known as source-tract interaction. Also, the aerodynamic noise source properties of amplitude and spectral content are directly affected by the vocal tract configuration because of changes in constriction areas and pressure distributions. Thus, the quasisteady assumption is suspect in certain phonetic environments. In fact, this has been a problem in using DFTs for the analysis of speech.

A multiresolution decomposition may help in this regard (Meyer, 1993). In such a decomposition, the high-frequency components can be more localized in time than the low-frequency components. Thus, the high-frequency components can change more rapidly than the low-frequency components without violating stationarity. Recent work has been done in multiresolution decomposition and its generalizations for analysis and compression of speech signals (e.g., Wickerhauser, 1993). These decompositions make use of wavelets, wavelet packets, and other orthonormal bases to find decompositions suitable for a given application. In the case of data compression Shannon entropy can be minimized (Coifman & Wickerhauser, 1993). For purposes of speech analysis, the multiresolution decompositions allow the analyst to tile the time-frequency plane tailored to the physical situation. While there is still a trade between frequency and time resolution because of the Heisenberg uncertainty principle, the duration of the time window can be tailored to the analysis frequency. Thus, a spectrogram would consist of time slices depending, not only on the time coordinate, but also the frequency coordinate. In the particular case of a wavelet transform, one obtains an octave-band decomposition. However, there are more general orthonormal bases that allow a more irregular tiling of the time-frequency plane, with a lower bound set on the area of a tile by the Heisenberg uncertainty principle.

It is proposed here that a multiresolution form of decomposition be used for articulatory synthesis, as well as, analysis. This would involve decomposing the time-dependent part of the equations of motion for air in the vocal tract into a general orthonormal basis. The vocal tract could still be divided into small tube sections for the spatial discretization, if desired. The matrices describing the transformation of pressure and volume velocities from one section to another (Sondhi and Schroeter's chain matrices) would be written in new orthonormal coordinates. In one possible implementation of a multiresolution synthesis area functions would be sampled at a fast rate, and this area function averaged over different intervals depending on the frequency scale of interest, with the higher frequencies requiring less duration for quasisteady conditions than the low frequencies. The time derivatives, including fractional derivatives, would be written in terms of the chosen orthonormal basis. While Fourier analysis transforms the derivative operator to a diagonal operator, the wavelet decomposition transforms the derivative into a sparse matrix. This sparse matrix can be used for fast computation of derivatives in the wavelet basis (Beylkin, 1993). Further, any noise sources can be can be shaped by bandpass filters composed of wavelets.

FOUR PARAMETER VOICE SOURCE

Articulatory synthesizers can have voice sources that are controlled using parameters other than articulatory parameters (e.g., Rubin, Baer, & Mermelstein, 1981). Or they have voice sources that are continuum mechanical models of the self-oscillating folds and air flow in the laryngeal region (Ishizaka & Flanagan, 1972). The latter simulations can require too much computation time or produce poor voice quality in running speech. While the former voice sources can produce natural sounding voice, they are not controlled by articulatory parameters.

The cover-body model of the vocal folds is the starting place for a four parameter model of the voice source (Hirano, 1974). It is supposed that all

aspects of voice quality having to do with solid structure can be determined by "...the relationship between the body and cover of the vocal cord" (Hirano, 1974, p. 91). The cover-body picture of phonation has been expanded by others, most notably Titze (1994), who has constructed muscle activation plots (MAPs) for fundamental frequency control. (While these MAPs have largely been based on canine data, Titze's group has recently measured stress-strain relations for the human vocal ligament (Titze, Min, & Alipour-Haghighi, 1994.) In these plots isofrequency contours are plotted against cricothyroid (CT) muscle activation and thyroarytenoid (TA) activation. However, for purposes of controlling the properties of the of the cover and body of the folds, the CT activation could be thought to represent any factor, intrinsic or extrinsic that controls the length of the cover and body, and stiffening both structures when they are lengthened. This change could be due to factors such as raising and lowering the larynx so that the larynx changes position along the spine, thus rotating the thyroid and cricoid relative to one another (Honda, 1995). Also, muscles who's primary effect is thought to be abductory and adductory motion of the folds can have an effect the length of the cover-body in ways analogous to the CT. On the other hand, TA activity, reduces the length of the cover-body complex, but by stiffening the body and relaxing the cover. There are many ways of attaining the same fundamental frequency using different combinations of CT and TA activation. However, these different combinations can often, if not always, be distinguished in other acoustic dimensions, including source amplitude and spectral content.

There are other ways to control fundamental frequency, source amplitude and spectral content. These include the degree of adduction and the transglottal pressure. The latter parameter is partly determined by what is happening in the upper vocal tract independent of the larynx. A tight constriction in the upper vocal tract and an open glottis will mean that transglottal pressure decreases. Thus, the four parameters in the proposed model of voice source control are the transglottal pressure, degree of abduction, (generalized) CT activity, and TA activity. These parameters should provide enough degrees of freedom to produce just about any voice quality. This would not be true if one of these parameters were omitted, and so this set could be considered minimal. Also, while it is not a detailed anatomical model of the larynx and its vibratory modes, it is sufficiently articulatory given the state of the art in articulatory synthesis.

REFERENCES

Beylkin, G. (1993). Wavelets and fast numerical algorithms. In I. Daubechies (Ed.), *Different perspectives on wavelets. Proceedings of Symposia in Applied Mathematics, Volume 47* (pp. 89-117). Providence: American Mathematical Society.

Bickley, C., Stevens, K. N., & Williams, D. R. (1994). A framework for synthesis of segments based on articulatory parameters. In *Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis.*

Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics, 20,* 27-38.

Coifman, R. (1991). Adapted multiresolution analysis, computation, signal processing, and operator theory. In *Proceedings of the International Congress of Mathematicians, Kyoto, 1990.* Tokyo: Springer-Verlag.

Coifman, R. R., & Wickerhauser, M. V. (1993). Wavelets and adapted waveform analysis. A toolkit for signal processing and numerical analysis. In I. Daubechies (Ed.), *Different perspectives on wavelets. Proceedings of symposia in applied mathematics, Volume 47* (pp. 119-153). Providence: American Mathematical Society. .

Davies, P. O. A. L., McGowan, R. S., & Shadle, C. H. (1993). Practical Flow Duct Acoustics. In I. R. Titze (Ed.), *Vocal fold physiology: Frontiers in basic science.* San Diego: Singular Publishing Group, Inc.

Guenther, F. H. (1994). *Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production.* (Technical Report CAS/CNS-94-012). Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems, Boston, MA.

Hirano, M. (1974). Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatrica, 26,* 89-94.

Honda, K. (1995). Laryngeal and extra-laryngeal mechanisms of F0 control. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues. For Katherine Safford Harris* (pp. 215-232). Woodbury, NY: AIP Press.

Ishizaka, K., & Flanagan, J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *The Bell System Technical Journal, 51.* 1233-1268.

Lin, Q. (1994). Vocal-tract computation: How to make it robust and fast. *Journal of the Acoustical Society of America, 96.* 2576-2579.

Maeda, S. (1982). A digital simulation method of the vocal-tract system. *Speech Communication, 1,* 199-229.

Mattingly, I. G. (1990). The global character of phonetic gestures. *Journal of Phonetics, 18,* 445-452.

McGowan, R. S., & Saltzman, E. L. (in press). Incorporating aerodynamic and laryngeal components into task dynamics. *Journal of Phonetics.*

Meyer, Y. (1993) *Wavelets, algorithms & applications.* Philadelphia, PA: Society for Industrial and Applied Mathematics.

Parthasarthy, S., & Coker, C. H. (1992). On automatic estimation of articulatory parameters in a text-to-speech system. *Computer Speech and Language, 6,* 37-75.

Rose, R. C., Schroeter, J., & Sondhi, M. M. (1994). An investigation of the potential role of speech production models in automatic speech recognition. In *Proceedings of the International Conference on Spoken Language Processing* (pp. 575-578). *September 18-22, Yokohama, Japan, Volume 2.*

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America, 93*, 1109-1121.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamic approach to gestural patterning in speech production. *Ecological Psychology, 14*. 333-382.

Schroeter, J., & Sondhi, M. M. (1992). Speech coding based on physiological models of speech production. In S. Furui & M. M. Sondhi (Eds.), *Advances in speech signal processing* (pp. 231-268). New York: Marcel Dekker.

Shirai, K., & Kobayashi, T. (1986). Estimating articulatory motion from speech wave. *Speech Communication, 5*, 379-385.

Sondhi, M. M., & Schroeter, J. (1987). A hybrid time-frequency domain articulatory speech synthesizer. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 35*, 955-967.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs: Prentice-Hall, Inc.

Titze, I. R., Min, Y. B., & Alipour-Haghighi, F. (1994). Stress-strain* response of the human vocal ligament and its effect on F0 control. *Journal of the Acoustical Society of America, 96*, 3324.

Wickerhauser, M. V. (1993). Best-adapted wavelet packet bases. In I. Daubechies (Ed.), *Different perspectives on wavelets. Proceedings of Symposia in Applied Mathematics, Volume 47* (pp. 155-171). Providence: American Mathematical Society.

FOOTNOTE

*Presentation of the invited organizer of the structured session on speech production at the International Congress of Acoustics, Trondheim, Norway, June 1995.

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 31-48

Nonsegmental Influences on Velum Movement Patterns: Syllables, Sentences, Stress, and Speaking Rate*

Rena A. Krakow†

I. INTRODUCTION

Investigations of the motor organization of speech show that if we can identify individual segmental requirements, we can begin to predict the manner in which segments will influence each other in fluent speech. That is, we can model coarticulation as the outcome of temporal overlap (coproduction) among characteristic speech movements for successive segments (e.g., Bell-Berti & Harris, 1981; Fowler, 1980; Munhall & Löfqvist, 1992; Saltzman & Munhall, 1989). Support for a coproduction model has largely been drawn from studies of articulators that shape the oral tract (the lips, jaw, tongue) or of formant frequencies that reflect oral tract shape (cf. Bell-Berti & Harris, 1981; Boyce, 1988; Fowler, 1980; Öhman, 1966; Saltzman & Munhall, 1989). However, recent work on the velum provides additional strong support for this framework (Bell-Berti & Krakow, 1991a).

Studies of velic movement patterns provide evidence for a segmental level of organization, with *n*-ary values of velic height ranging from extreme low to extreme high positions. These differences in intrinsic velic height are attributable not only to whether a segment is nasal or not, and whether it is a vowel or a consonant, but to such factors as vowel height, consonant place, manner, and voicing (cf. Bell-Berti, 1980, 1993; Bell-Berti & Hirose, 1975; Clumeck, 1976; Henderson, 1984; Mattisoff, 1975; Moll, 1965; Ohala, 1971, 1975; Ushijima & Sawashima, 1972). Velic gestural patterns in sequences of segments also show that there is temporal overlap among the characteristic movements for adjacent segments (Bell-Berti, 1980; Bell-Berti & Krakow, 1991a).

In a number of studies of velum movement, a failure to recognize that oral segments could vary in their intrinsic velic requirements resulted in the identification of the earliest onset of velic lowering in a CV_nN sequence (C=an oral consonant; V_n = any number of phonemically oral vowels; and N = a nasal consonant) as the onset of velic coarticulation for the nasal consonant (e.g., Benguerel, Hirose, Sawashima, & Ushijima, 1977; Bladon & Al-Bamerni, 1982; Kent, Carney, & Severeid, 1974; Moll & Daniloff, 1971). However, further research showed that velic lowering is also observed in the transition from a consonant to a vowel in phonemically oral CV_nC sequences, because oral vowels have intrinsically lower velic positions than oral consonants (e.g., Bell-Berti, 1980; Clumeck, 1976; Henderson, 1984; Ushijima & Sawashima, 1972). Still, it is often the case that velic lowering for a phonemically oral vowel or vowels in a CV_nN sequence is temporally overlapped with the larger lowering gesture for the nasal consonant, rendering the discrete movement components indistinguishable. Adding time between the oral consonant and the nasal consonant, by inserting additional vocalic segments and/or by slowing the rate of speech, reduces the overlap between the shallow velic lowering gesture for the vowel(s) and the more extreme lowering for the nasal consonant (Bell-Berti & Krakow, 1991a).

Identification of a segmental level of gestural organization and an understanding of the manner in which gestures for successive segments combine provide only a partial understanding of the nature of speech motor organization, however, because

The research reported in this chapter was supported by NIH Grants DC-00121 and HD-01994 to Haskins Laboratories. I thank Ignatius Mattingly and Fredericka Bell-Berti for comments on an earlier draft of the article and for many stimulating discussions on speech motor organization and the role of the velum.

there are a number of other sources of influence. These include (but are not limited to) variations in syllable structure, syllable location in a phrase or sentence, stress, and speaking rate. Hence, a more complete model of speech motor organization would specify the various nonsegmental as well as segmental influences on the articulators (cf. Fujimura, 1990; Kent & Minifie, 1977; Macchi, 1988; Nittrouer, Munhall, Kelso, Tuller, & Harris, 1988; Vaissière, 1988). Research on the velum shows that the nonsegmental influences are robust and provide additional evidence that the phonetic and phonological functions of the velum are varied and important (Bell-Berti & Krakow, 1991a; Fujimura, 1990; Krakow, 1987, 1989; Vaissière, 1988).

A number of specific examples help to illustrate this point: First, patterns of coordination among velic and labial gestures distinguish syllables with final, from those with initial, bilabial nasal consonants (Krakow, 1989). Second, velic movement patterns provide support for the notion that declination over the course of a phrase or sentence is not limited to laryngeal-respiratory behavior; similar patterns to those reported for F0 and acoustic amplitude have been found for the jaw, vowel formant frequencies, and for the velum (cf. Bell-Berti & Krakow, 1991b; Gelfer, 1987; Krakow, Bell-Berti, & Wang, 1991; Vatikiotis-Bateson & Fowler, 1988; Vayra & Fowler, 1992). Third, velic movements provide support for the hypothesis raised by Schourup (1973), based on cross-language phonological data, that stress enhances the likelihood of assimilatory nasalization on low vowels (Krakow, 1987; Vaissière, 1988). And fourth, velic movement patterns add to our knowledge about how speakers reorganize their gestures to produce speech at a faster rate (Bell-Berti & Krakow, 1991a; Kent et al., 1974; Kuehn, 1976).

This article focuses on nonsegmental influences (syllable organization, syllable position in a sentence, stress, and speaking rate) on velic movements, describing them in terms of characteristic articulatory patterns. The data also indicate a need to incorporate the notion of variable strategies in speech production models, since the evidence suggests that speakers may vary in the manner in which they implement some of these nonsegmental changes (cf. Kent et al., 1974; Kuehn, 1976; Vaissière, 1988). The research described here, combined with studies on the segmental organization of velic gestures (described in Bell-Berti, 1993), call for a much enriched model of velic control (see also Bell-Berti, 1980; Bell-Berti, 1993; Fujimura, 1990; Vassiere, 1988).

II. SYLLABLES

The notion that syllables are units of motor organization goes back at least to 1928, when Stetson proposed that pulses of expiratory muscle activity divide the speech stream into syllable-sized units. Stetson's own data on pulmonary air pressure and his observations of chest wall movement appeared to support this hypothesis (see Stetson, 1951). However, subsequent electromyographic research by Draper, Ladefoged, and Whitteridge (1960) showed that there was no systematic relation between respiratory muscle activity and the syllable. In later studies, the focus shifted from respiratory to articulatory patterns that might correlate with syllable organization. For example, Kozhevnikov and Chistovich (1965) proposed that the articulatory syllable was coextensive with the domain of anticipatory coarticulation and studies of lip protrusion activity provided preliminary support for this hypothesis (e.g., Daniloff & Moll, 1968; Kozhevnikov & Chistovich, 1965; Tatham, 1970). But subsequent work, showing asynchronous coarticulatory domains for different articulators, established the fundamental flaw of this hypothesis (cf. Kent et al., 1974; Kent & Minifie, 1977; Öhman, 1966). That the proposed syllable-based temporal patterns were not generalizable across articulators prompted the question of whether the regularities might rather be found in the patterns of coordination among the articulators (see Krakow, 1989), a suggestion that is consistent with current approaches to characterizing linguistic units in terms of stable patterns of gestural coordination (Browman & Goldstein, 1988; Macchi, 1988; Nittrouer et al., 1988).

The timing of velic gestures relative to gestures that shape the oral tract provides a particularly interesting domain in which to seek stable syllable-based patterns for several reasons: First, the phonetic and phonological evidence suggests that patterns of velic lowering and nasal assimilation are affected by the position of a nasal consonant in a word and hence, possibly in a syllable. Word-final nasal consonants are produced with greater and earlier velic lowering than word-initial nasal consonants, and correspondingly, nasal assimilation is more likely to affect vowels preceding word-final nasals than those preceding or following word-initial nasals (cf. Clumeck, 1976; Fujimura, 1990; Fujimura, Miller, & Kiritani, 1975, 1977; Henderson, 1984;

Ohala, 1971; Schourup, 1973; Vaissière, 1988). Second, a conflict between studies showing that word boundaries inhibit anticipatory velic lowering and studies showing no word boundary effect can be resolved in favor of an argument for syllable-based patterning (see Krakow, 1989). That is, the studies claiming to find a word boundary effect compared sequences of the forms CV#NV and CVN#V (e.g., J. Ohala, 1971; M. Ohala, 1975), and those that found no effect compared forms such as CVVN and CV#VN (e.g., Moll & Daniloff, 1971). If syllables, rather than words, are relevant and if anticipatory velic lowering is more extensive for syllable-final than syllable-initial nasals, then one would expect to see precisely the patterns reported: that is, less of an effect on the vowel before the nasal consonant in CV#NV than CVN#V (where a sequence with a syllable-initial nasal is compared to one with a syllable-final nasal) but no difference between CVVN and CV#VN (where two sequences with syllable-final nasals are compared). Third, because the velum functions as an independent articulatory subsystem (see Browman & Goldstein, 1986), the velic lowering gesture for a nasal consonant can, at least theoretically, shift in time relative to the tongue (for /n/) or lip gesture (for /m/) to signal different syllable organizations.

The hypothesis that oral-velic timing patterns for nasal consonants would reveal syllable organization was explored by Krakow (1989). Table 1 shows the utterance list that was used to investigate syllable- and word-based patterns of labio-velic coordination for bilabial nasal consonants.[1] The stimuli were designed to control segmental influences while providing for the following comparisons: word-initial vs. word-final nasals (columns 2 and 3); syllable-final nasals in word-final vs. word-medial positions (columns 4 and 5); word-medial nasals of unclear syllable affiliation vs. word-initial and word-final nasals (column 1 vs. columns 2 and 3). For the purposes of this study, no a priori assumptions were made about the syllable affiliation of a nasal consonant in word-medial intervocalic position (or in a CV\|NVC sequence).

Two speakers produced 12 repetitions of each sequence in a brief carrier phrase. Vertical velic movements were tracked with the Velotrace (Horiguchi & Bell-Berti, 1987) and an optoelectronic tracking system (see Krakow & Huffman, 1993). The tracking system was also used to monitor the time-varying vertical lower lip position, which is influenced by both lower lip and jaw activity. The movement traces were aligned with reference to the onset of bilabial contact for the nasal consonant, a measure derived with the use of an Electrolabiograph.[2]

Table 1. *Stimuli for experiment on syllables.*

NASAL CONSONANT POSITION				
1	2	3	4	5
Word-Medial	Word-Initial	Word-Final	Word-Final	Word-Medial
homey	hoe me	home E	home Lee	homely
Seymour	see more	seam ore	seam lore	
seamy	see me	seam E	seam Lee	seemly
helmet	hell mitt	helm it	hem lit	hemlette
pomade	pa made	palm aid		

Movement onsets and offsets of the lower lip and velum were determined with the use of the corresponding instantaneous velocity traces. A noiseband around zero velocity (determined for each subject and each articulator) was used for the purpose of eliminating from the measured movements those portions of the movement trajectories that appeared as drift, with velocity values hovering around zero (Krakow, 1989).

Figure 1 provides sample movement patterns for comparison pairs with word-initial vs. word-final nasal consonants. Data from different phonetic sequences and different subjects are used to illustrate the stability of the gestural differences in the two conditions. The vertical line in the middle of each panel marks the onset of bilabial contact. (Lip raising following bilabial contact represents compression of the lower lip against the upper lip.) As shown, the lip movements associated with the bilabial nasal were quite similar for the corresponding word-inital and word-final nasals. In contrast, the velic movements were remarkably different across the word-position manipulation; most obvious, perhaps, is the presence of the long low plateau in the sequences with final nasals, and the absence of such a plateau in the words with initial nasals. The velum typically reached a lower minimum position for the word-final nasals as well. Considering the coordination between the lip and velum, it can be seen that, regardless of the effects of segmental context and/or speaker, there were two distinct and stable patterns, one associated with word-initial nasals, and the other, with word-final nasals. That is, the achievement of the velic target co-occurred with completion of lip raising for initial nasals, but with initiation of lip raising for final nasals. As a result of this dif-

ference, the vowel preceding the final nasal was associated with considerably lower velic height than the vowel preceding the initial nasal. This pattern is clearly consistent with phonological evidence that vowels before word-final nasals are more likely to be nasalized than are vowels before word-initial nasals.

To see whether the patterns described reflected syllable organization, word organization, or some combination, labio-velic movement patterns were compared for sequences having syllable-final nasals in word-medial and word-final positions. Figure 2 provides sample comparisons for the two subjects' productions of *home Lee* vs. *homely*, and *seam Lee* vs. *seemly*. The patterns bear a clear resemblance to those shown in Figure 1 for *home 'E'* and *seam 'E'*: A low velic plateau is evident in all sequences with syllable-final nasals, although the plateau is somewhat shorter for those nasals in word-medial position, reflecting the shorter vocalic duration of the first syllable of the bisyllabic single word than the first word of the matched bisyllabic two-word sequence. Nonetheless, velic lowering offset and lip raising onset continued to be temporally linked for syllable-final nasals, despite the change in word position. This inter-articulator pattern was therefore taken to be an indicator of the presence of a syllable-final nasal consonant.

Similarly, the pattern described for word-initial nasals was found to be syllable-based. The pair, *pygmy—pig me*, was used by Krakow (1989) to ensure a syllable break before the word-medial /m/ to contrast a word-medial syllable-initial nasal and a word-initial nasal. The sequence /gm/ is not, however, compatible with any alternative placement of the syllable boundary, limiting its use to this comparison alone. Velum lowering offset was timed to lip raising offset in both of these sequences for the two subjects, consistent with the pattern described for the word-initial nasals above.

It is possible, therefore, to describe a bi-stable pattern of labio-velic coordination that distinguishes syllables with initial bilabial nasal consonants from those with final nasal consonants: The end of velum lowering aligns either with the beginning or the end of lip raising toward the position required for the bilabial nasal consonant.

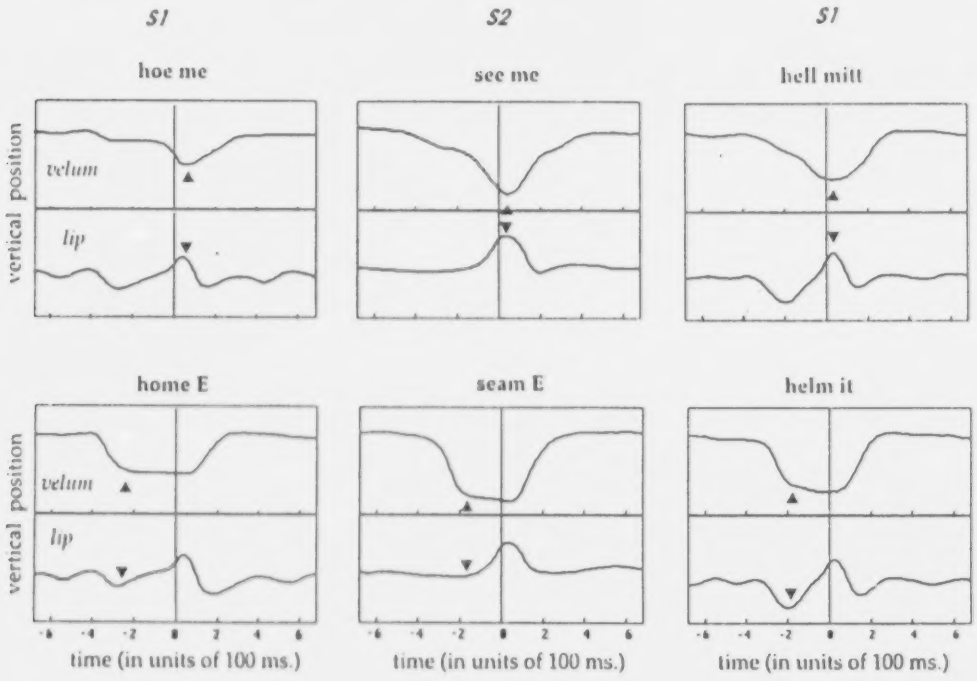


Figure 1. **Sample velum and lower lip movements (in the form of ensemble averages) for sequences with word-initial (top panels) and word-final (bottom panels) nasal consonants from S1 and S2. The vertical line in the middle of each panel marks the onset of bilabial contact for the /m/. The triangles in the panels mark velum lowering offset and the coordinated event in the lower lip movement.**

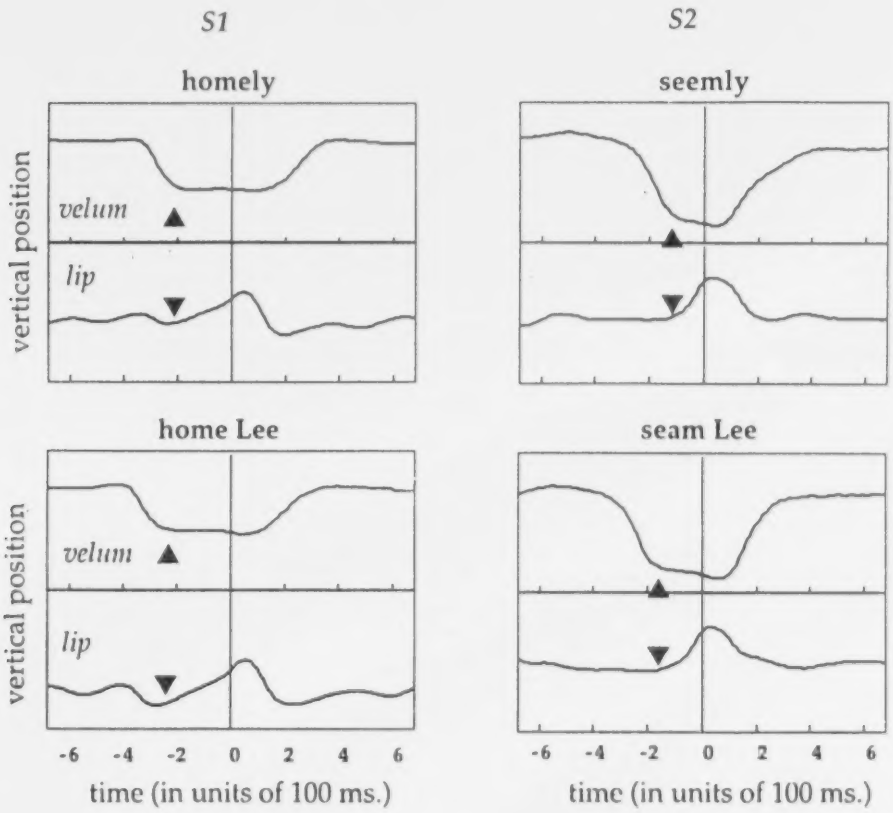


Figure 2. **Sample velum and lower lip movements (in the form of ensemble averages) for sequences with syllable-final nasals in word-medial (top panels) and word-final positions (bottom panels) from S1 and S2. The vertical line in the middle of each panel marks the onset of bilabial contact for the /m/. The triangles in the panels mark velum lowering offset and the coordinated event in the lower lip movement.**

The distinct patterns of labio-velic coordination observed for syllable-initial vs. syllable-final nasals made it possible to compare the gestural patterns for word-medial nasals whose affiliation is unclear, i.e., phonotactic constraints do not decide between syllabification with the preceding vs. the following vowel (Table 1, column 1). It has been proposed that, in some instances, such a consonant might be "ambisyllabic," i.e., a member of both syllables (For further discussion of ambisyllabicity, see Kahn, 1976; Selkirk, 1982). However, the movement patterns for the sequences in Column 1 indicated that the nasal consonant was typically affiliated with the preceding or the following vowel, but not simultaneously with both. In most cases, syllabification was determined by the stress pattern, with primary stressed syllables attracting the nasal consonant. Thus, for example, the nasal consonants in *seamy* and *homey* appeared to be syllable final (velic lowering offset timed to lip raising onset), whereas the nasal consonant in *pomade* appeared to be syllable initial (velic lowering offset timed to lip raising offset), for both speakers. Figure 3 provides sample movement patterns showing stress-based syllabification of medial nasals.

The remaining two sequences of this type, *helmet* and *Seymour,* showed more variable patterning. The movement patterns of one subject (S1) showed syllable-final organization of the nasal consonant for both of these words, linking the nasal with the primary stressed syllable. However, the patterns of the second subject (S2) varied from token to token for these sequences. Some tokens of *helmet* showed syllable-final organization of the nasal consonant, and some showed syllable-initial organization (Figure 4). Similarly, some tokens of *Seymour* showed syllable-final, and some, syllable-initial, organization of the nasal consonant. Still other tokens of *Seymour* were hard to characterize as one or the other type.

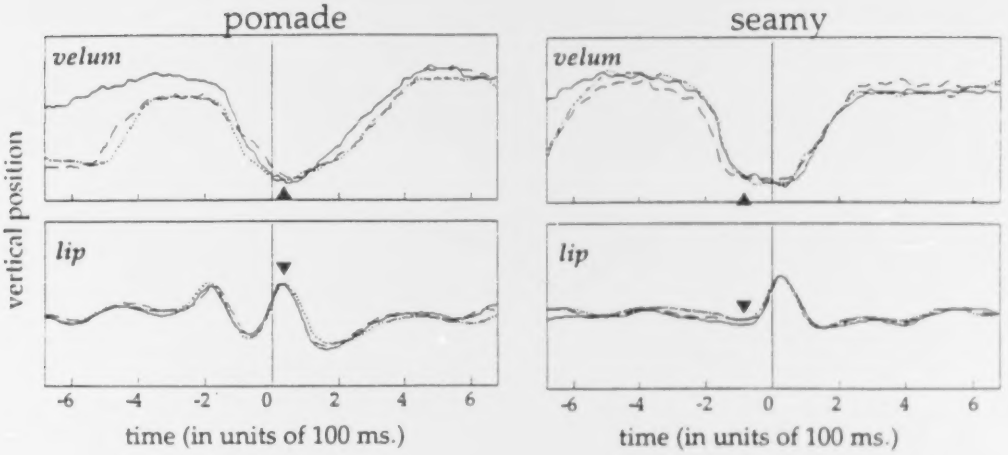


Figure 3. **Sample velum and lower lip movements for sequences with word-medial intervocalic nasals, from S2. Three tokens are displayed in each panel with different line types. The vertical lines in the panels mark the onset of bilabial contact for the /m/. The triangles mark velum lowering offset and the coordinated event in the lower lip movement. The nasal consonant in *pomade* shows syllable-initial organization whereas the nasal consonant in *seamy* shows syllable-final organization.**

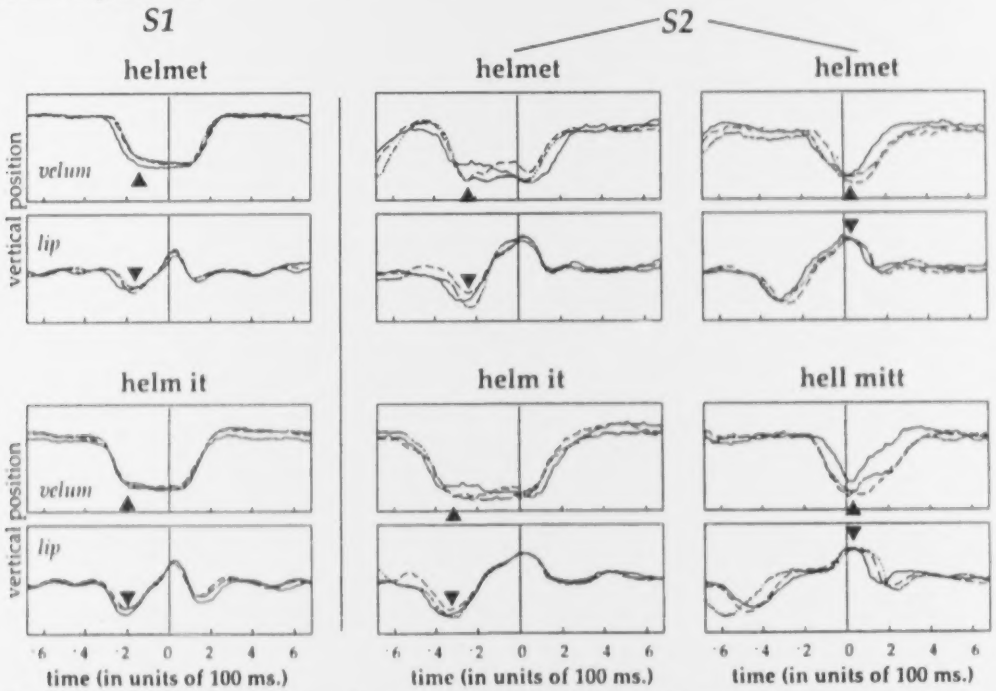


Figure 4. **Labio-velic coordinative patterns for tokens of *helmet (top panels)* shown above tokens of *helm it* or *hell mitt*, depending upon whether the patterns for *helmet* resemble those for *helm it* or *hell mitt*. The vertical line in the middle of each panel marks the onset of bilabial contact for the /m/. The triangles mark velum lowering offset and the coordinated event in the lower lip movement. Three tokens are displayed in each panel. S1 had a consistent pattern while S2 alternated between two patterns.**

It seems clear that the question of ambisyllabicity and of variation in the production of sequences of this sort requires a larger data set, with additional subjects, sequences, and tokens. However, some preliminary observations can be made. First, despite some evidence of variability in patterning, the vast majority of the tokens in this set (column 1, Table 1) were organized such that the nasal consonant appeared syllable-final or syllable-initial, but not a combination. And, generally the nasal consonant affiliated with the syllable that received primary stress. Hence, these data provide little support for the notion of ambisyllabicity as a combination of syllable-initial and syllable-final attributes but the data do suggest that some word-medial consonants are likely to show free variation with respect to syllable affiliation. Considering the two that varied here, i.e., the nasal consonants in *helmet* and *Seymour,* the following observations can be made: A speaker may affiliate a medial consonant with the upcoming syllable instead of the preceding syllable which bears primary stress, (a) when the stressed syllable already contains a post-vocalic consonant, as in the case of *helmet,* or (b) when the following syllable receives secondary stress, as in the case of *Seymour.* Further study is necessary to determine, with greater certainty, the kinds of sequences that are likely to vary in this way.

Returning, however, to cases in which the nasal consonant is unambiguously syllabified (e.g., those in Columns 2, 3, 4, and 5 of Table 1), much stronger claims can be made. For such sequences, there are characteristic patterns of inter-articulator timing associated with bilabial nasal consonants in different syllable positions. A consistent pattern of bi-stability distinguished sequences with syllable-initial vs. syllable-final nasals: the end of velic lowering coincided with the end of lip raising for the syllable-initial nasals (whether word-medial or -initial), but with the beginning of lip raising for the syllable-final nasals (whether word-medial or -final). These results strongly support the notion of the syllable as an articulatory unit, and suggest as well that additional insight into the question of ambisyllabicity is likely to be found in speech production data.

III. SENTENCES

Recent work on velic movements shows that, in addition to the effects of a segment's position in a syllable, there are effects of a segment's position in larger domains—such as a phrase or sentence (the span of a breath group). Over the course of such domains, one observes an overall decline in peak velic position for obstruents (Bell-Berti & Krakow, 1991b; Krakow, Bell-Berti, & Wang, 1991). This pattern is similar to the declination of fundamental frequency over the course of a phrase or sentence, long noted in the phonetics literature (see Breckenridge, 1977; Cooper & Sorenson, 1981; Gelfer, 1987; Maeda, 1976). While F0 declination has been treated by some as resulting from intentional control of muscles affecting F0 (Breckenridge, 1977; Cooper & Sorenson, 1981), Gelfer's (1987) acoustic and physiological research shows that F0 declination, along with a concomitant decline in acoustic amplitude, is largely the passive consequence of declining subglottal pressure over the course of a sentence.

Recent findings of declination in measures of the jaw (Vatikiotis-Bateson & Fowler, 1988; Vayra & Fowler, 1992), the velum (Bell-Berti & Krakow, 1991b), and vowel formant-frequency patterns (reflecting tongue-jaw positions) (Vayra & Fowler, 1992) lend support to the notion that declination is a general phenomenon, with effects on laryngeal-respiratory and supralaryngeal events in the production of spoken sentences (Fowler, 1980). For example, kinematic data collected by Vatikiotis-Bateson and Fowler (1988) showed weak, but consistent, declination in the extent, peak velocity, and duration of jaw opening and closing gestures over the course of reiterant speech utterances. Vayra and Fowler (1992) examined measures of F1 and F2 for the vowels in isolated trisyllabic Italian pseudowords. The results showed that the vowels became more centralized from early to late: open vowels became less open (F1 decreased) and closed vowels became less closed (F1 increased), results consistent with increased relaxation or reduced energy later in the utterance. The F2 measures indicated that there was also centralization along the front-back dimension.

These results, along with informal observations of what appeared to be declination patterns in velic data collected for other purposes, prompted a systematic investigation of possible velic declination (Bell-Berti & Krakow, 1991b; Krakow, Bell-Berti, & Wang, 1991). The set of seven utterances constructed for the study, ranging in length from three to nine syllables, is listed in Table 2. Each natural sequence was paired with a reiterant sequence consisting of the appropriate number of repetitions of the syllable *ten*. The natural sentence was produced first, followed by, and functioning as a model for, the reiterant version. The natural sequences were designed to minimize phonetic context effects on the velum

(and still be meaningful); the reiterant sequences were designed to eliminate such effects entirely. Each of the subjects produced 12 randomized repetitions of each natural-reiterant sentence pair. Velum movement was monitored with the Velotrace and a modified Selspot system. Measurements were made of peak velic positions for each /t/ in the reiterant sentences and for the first and last /s/ in the natural sentences (i.e., the /s/ in *Sue* and the /s/ in *Sid*).

Table 2. *Stimuli for velum declination experiment.*

Natural Sentences	Reiterant Sentences
Sue saw Sid.	Ten ten ten.
Suzy saw Sid.	Tenten ten ten.
Suzy saw sad Sid.	Tenten ten ten ten.
Suzy saw sexy Sid.	Tenten ten tenten ten.
Suzy saw sad sexy Sid.	Tenten ten ten tenten ten.
Suzy saw sexy sassy Sid.	Tenten ten tenten tenten ten.
Suzy saw sad sexy sassy Sid.	Tenten ten ten tenten tenten ten.

Figure 5 provides sample movement patterns for some of the reiterant sequences, showing that the velum is consistently lower at the last, than at the first, peak of each sentence. A lower velic peak at the end than at the beginning of the sentences was observed for all sequence types (whether natural or reiterant, short or long) and for all subjects. A reduction in the height of velic peaks over the course of a sentence is consistent with a notion of decreasing energy over its time course, as higher velic positions are normally associated with increased activity of the levator palatini (see Bell-Berti, 1993, for a review of muscle activity in velopharyngeal function).[3] The patterns shown in Figure 5, however, indicate that while there was a general decline, there were also some local increases and decreases, similar to patterns observed for F0 and acoustic amplitude (Gelfer, 1987).

To determine the nature of such local effects on the velic movements, it was necessary to examine changes in velic measures in the middle of the utterances and to control for possible stress effects or effects of syllable position within a word. For this purpose, the reiterant syllables corresponding to the bisyllabic words (*Suzy, sexy,* and *sassy*) in the natural sequences were examined. Peak velic position was obtained for the reiterant syllables corresponding to the stressed and unstressed syllables in each bisyllabic word. In this way, it was possible to compare velic peaks for stressed syllables occurring earlier and later, and for unstressed syllables occurring earlier and later. Comparisons were made in a pair-wise fashion because some sentences had only two of the three bisyllabic words (see Table 2).

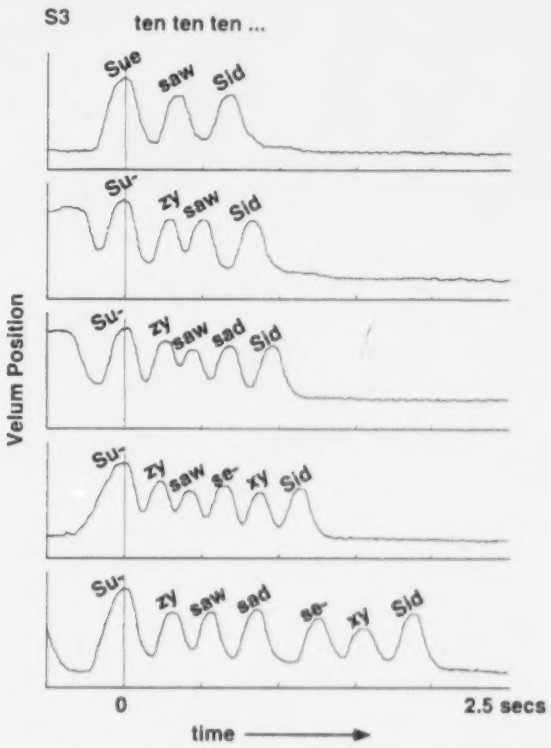


Figure 5. **Sample velum movements for tokens of reiterant sentences of varying length from a subject in the declination study. The natural sentences that served as models appear above the movements for the reiterant sentences. Each peak in the movements is associated with a /t/ from one of the reiterant syllables.**

The mean peak velic positions in the reiterant syllables corresponding to the bisyllabic word pairs are plotted in the panels of Figure 6 for each subject. The measures were obtained only from those sentences which contained both members of a given comparison pair. The earlier word of the pair occurs at the left of each panel and the later word at the right. The following effects were observed. First, there was a consistent decline in peak velic position from earlier to later stressed syllables. Second, more often than not, unstressed syllables also showed a decline, but when present, it was always weaker than the decline in the stressed syllables.

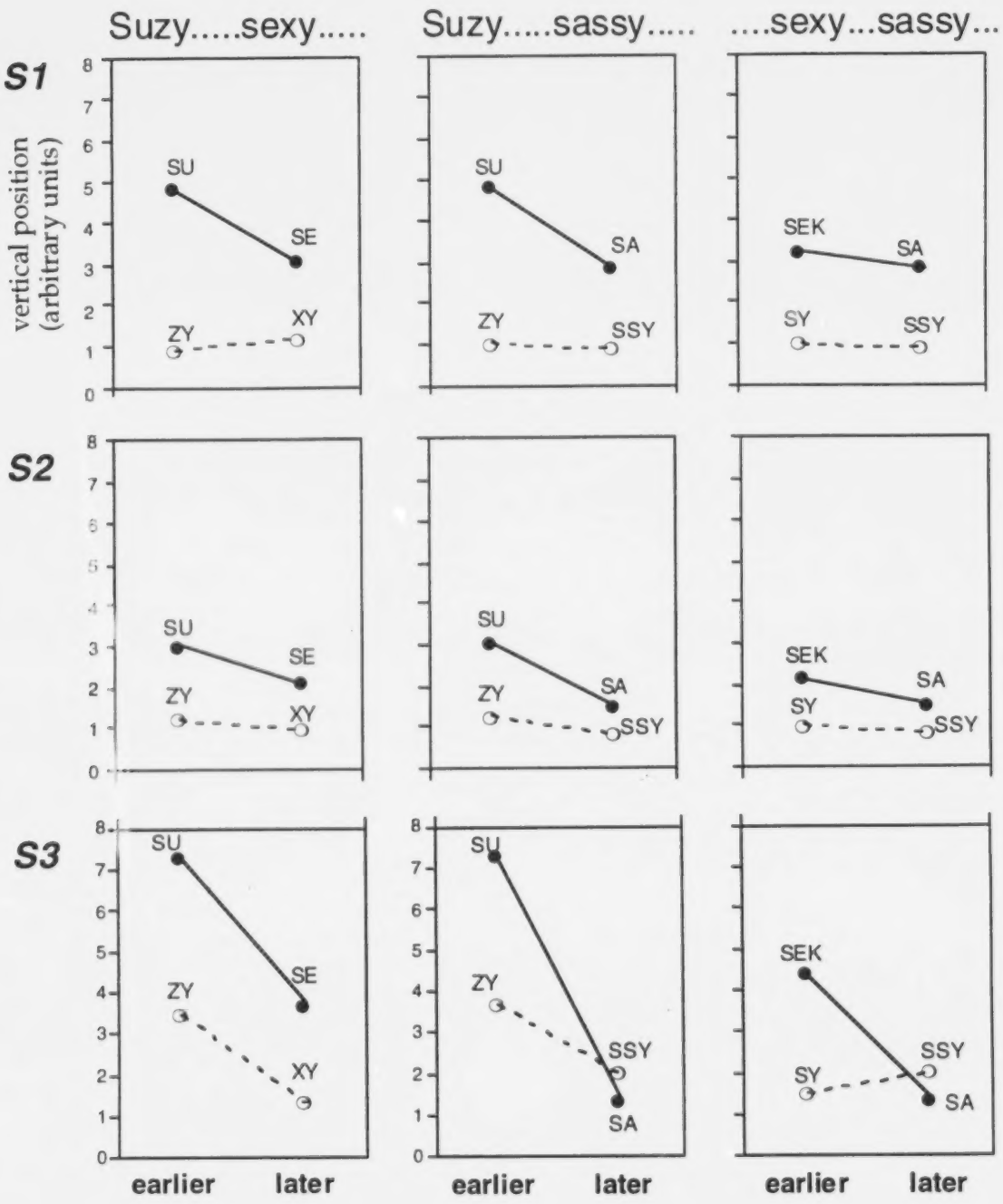


Figure 6. **Means for peak velum position associated with the /t/ of the reiterant syllables in earlier and later positions in the sentences. "Earlier" and "later" measures are, respectively, at the left and right of each panel. Panels at the left compare peaks in reiterant versions of *suzy* and *sexy*, those in the middle compare peaks in reiterant versions of *suzy* and *sassy*, and those at the right compare peaks in reiterant versions of *sexy* and *sassy*. Measures of stressed syllables are represented with filled circles, measures of unstressed syllables, with open circles. Data are shown for the three subjects.**

Stronger declination effects for stressed syllables were also reported by Vayra and Fowler (1992), based on their measures of jaw position and of F1 formant values for the vowel /a/ produced by Italian speakers.

Third, in all but two comparisons between earlier and later means, velic peaks were higher for stressed than for unstressed syllables. Whether these effects are attributable to stress or to within-word declination, or to a combination of the two, is unclear because stressed syllables were always first syllables and unstressed syllables were always second syllables. However, there are data that indicate that stress probably played an important role in these velic patterns. For example, Vaissière (1988) showed that velic peaks for the oral consonants in CVN sequences were higher in stressed than in unstressed syllables—just the pattern that was observed in the velic declination data. There is also evidence to suggest that supralaryngeal declination is largely a phrasal, rather than a word-level phenomenon. According to Vayra and Fowler (1992), declination patterns of the jaw and formant frequencies observed in isolated trisyllables are weakened or disappear when the trisyllables are embedded in carrier phrases. Given Vayra and Fowler's results and those of Vaissière (1988), the present findings suggest that there is a reduction in the contrast between velic peaks in stressed and unstressed syllables as they occur later in a sentence.

In general, there is evidence of declination of velic height over the course of a sentence from the first to the last syllable of a sentence and also over the medial regions of a sentence, with local perturbations due to stress, and possibly also to syllable position in a word. Taken together with the recent studies of formant frequency and jaw declination, the velic data provide support for the notion of declination as a more general phenomenon in speech production than was previously assumed. To date, declination patterns have been observed in respiratory-laryngeal behavior, in velic and mandibular movements, and in vowel formant frequencies. All are compatible with a notion of decreasing energy later in a sentence.

IV. STRESS

The preceding section prompts the question of whether there are manifestations of stress on velic movement patterns, apart from the higher peaks observed for oral consonants in stressed syllables (Vaissière, 1988). In his cross-language survey of nasalization, Schourup (1973) reported a relation between stress and nasalization: stressed vowels appeared more likely than unstressed vowels to become nasalized by assimilation. Although very few experimental studies have addressed stress effects on the velum, the results appear to be generally consistent with Schourup's report. Vaissière's (1988) subjects showed not only higher velic peaks for oral consonants in stressed syllables, but also lower velic valleys for nasal consonants in the stressed syllables. And Krakow's (1987, 1989) subjects kept the velum lowered for a longer time in stressed than in unstressed syllables beginning or ending with a nasal consonant.

Schourup was also interested in the relation between vowel height and nasalization, noting like others before and after him (e.g., Beddor, 1982; Chen, 1975; Henderson, 1984; Lightner, 1970) that low vowels are more likely to become nasalized than high or mid vowels. Taking the vowel height patterns together with his observations on the effects of stress, Schourup concluded that stressed low vowels are particularly susceptible to contextual nasalization. This raises the question of whether velic movements for all vowels are similarly affected by stress, a question that has not been systematically addressed in any of the studies in the literature.

Vaissière's work shows that for consonants (at least for obstruents and nasals), adding stress enhances intrinsic velic positional differences: the high velic positions associated with obstruents become higher, whereas the low velic positions associated with nasal consonants become lower. Although vowels tend to have velic positions intermediate between those of obstruents and nasal consonants, high vowels have intrinsically higher velic positions than low vowels. The difference is so robust that it affects velic positions for adjacent obstruents or nasal consonants (Bell-Berti, Baer, Harris, & Niimi, 1981; Henderson, 1984). Thus, it is possible that adding stress will merely enhance the intrinsic vocalic contrast by raising velic positions for high vowels, but lowering them for low vowels.

To address this question, Velotrace data were collected as two subjects produced 12 tokens of the sequences—/mábab/, /mabáb/, /míbab/, /mibáb/, /bábam/, /babám/, /bábim/, /babím/—in a brief carrier phrase. This set of sequences provided comparisons between velic positions for stressed and unstressed /i/ and /a/ following an initial nasal and preceding a final nasal. Positional measures, designed to span the duration of each vowel adjacent to a nasal consonant, were obtained for a number of acoustically-based reference points: for words with

initial nasal consonants, velic height was measured at first vowel onset, midpoint, and offset; for words with final nasals, velic height was measured at second vowel onset, midpoint, and offset.

The results for syllables with initial nasals are shown in Figure 7, which displays, for each subject, the velic height measures obtained for the three measurement points averaged over the twelve tokens of each subject's productions. Intrinsic differences in velic height between /ɑ/ and /i/ were evident for both subjects in stressed and unstressed syllables: the velum was lower for /ɑ/ than for /i/ at all corresponding measurement positions. For S1, stress enhanced these intrinsic differences: In syllables beginning with a nasal consonant, the velum was higher for /i/ when stressed than when unstressed, but lower for /ɑ/ when stressed than when unstressed, for most of the vowel's duration. For S2, however, the velum was lower for both stressed /ɑ/'s and stressed /i/'s following the initial nasal than for unstressed /ɑ/'s and /i/'s, respectively.

Figure 8 shows the corresponding data for vowels preceding final nasals in /bɑm/ and /bim/. Again, it can be seen that there were intrinsic differences in velic height between the two vowels, with lower positions for /ɑ/ than for /i/; for these sequences, the differences were evident at the vowel midpoint and offset for both subjects. And, as with the vowels following a nasal consonant, stress effects on vowels preceding a nasal consonant were different for the two subjects. For S1, the velum was higher for stressed than for unstressed /i/, but lower for stressed than for unstressed /ɑ/. For S2, however, the velum was lower in stressed than in unstressed syllables for both /ɑ/ and /i/.

Clearly, the two subjects used two different strategies for stressing syllables, at least with respect to velic movements. For one subject, stress enhanced the already existing contrast between the high and low vowels by increasing velic height in the former case and decreasing height in the latter. For the other subject, stress resulted in a lower velum, regardless of vowel height.

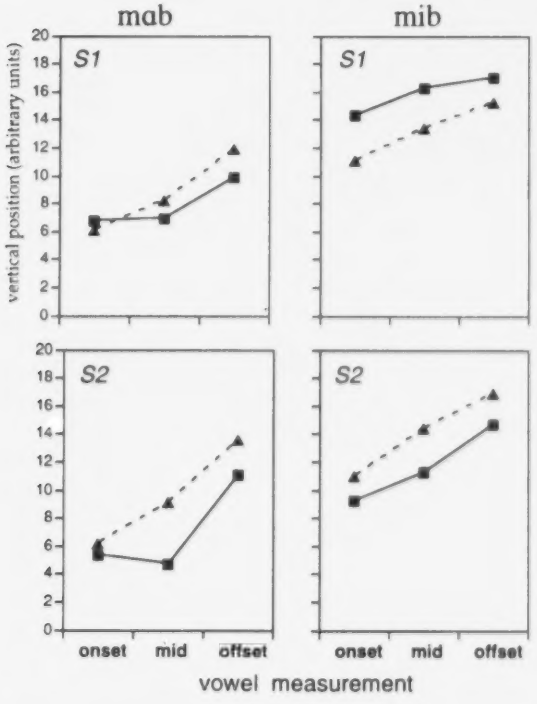


Figure 7. **Mean velum position at the onset, the middle, and the offset of stressed (squares) and unstressed (triangles) vowels following an initial nasal consonant. The vowels /ɑ/ and /i/ are represented in the left and right panels, respectively. Data are shown for the two subjects.**

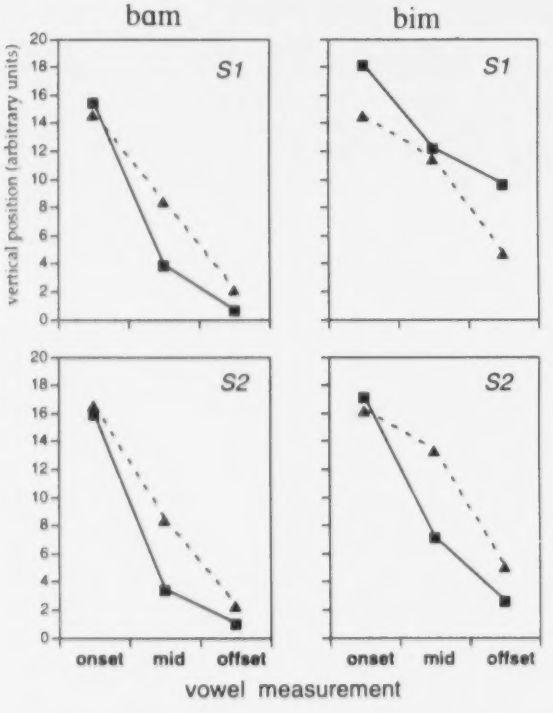


Figure 8. **Mean velum position at the onset, the middle, and the offset of stressed (squares) and unstressed (triangles) vowels preceding a final nasal consonant. The vowels /ɑ/ and /i/ are represented in the left and right panels, respectively. Data are shown for the two subjects.**

BEST COPY AVAILABLE

Similar variability in the effects of stress has been found for the tongue. For example, Kent and Netsell (1971) found that stress enhanced the intrinsic differences in tongue position for different segments for their two speakers. That is, /i/ had a higher—fronter tongue position and /æ/ had a lower-fronter tongue position when stressed than when unstressed. These data parallel the present findings for S1. In contrast, Houde (1967) reported that, for his speaker, the tongue positions for high and low vowels (/i/, /u/, /ɑ/) were consistently lower in stressed than in unstressed syllables. These data parallel the current findings for S2. It would be interesting to know whether the two subjects whose velic movements were examined here and showed differing effects of stress would show parallel differences in their tongue movements.

As for Schourup's claim that stress increases the likelihood of assimilatory nasalization on vowels, the present data suggest that this may not hold for both high and low vowels for all speakers. However, a more specific claim of Schourup that vowels most likely to be contextually nasalized are low, stressed, and precede a final nasal consonant does find support in these data. For both subjects in this study, the velum was lowest for stressed /ɑ/'s preceding a final nasal consonant (cf. Figures 7 and 8).

V. SPEAKING RATE

Another nonsegmental variable affecting velic movement patterns is speaking rate. Studies of the velum (Bell-Berti & Krakow, 1991a; Moll & Shriner, 1967), like those of other articulators (cf. Browman & Goldstein, 1991; Hardcastle, 1985; Munhall & Löfqvist, 1992; Saltzman & Munhall, 1989), indicate that as the rate increases, there is greater overlap among articulatory gestures for adjacent and near-adjacent segments. For example, Bell-Berti and Krakow (1991a) showed that, at a speaker's self-selected normal rate, the velic movement pattern for a sequence like *"It's **a lan**sal again"* contains a small lowering movement from the peak for the obstruent /s/ of /ɪts/ toward a somewhat lower position for the /əlɑ/ sequence, followed by a more extreme lowering movement for the nasal consonant. At a rapid rate, however, the small movement for the vocalic portion (including the /l/, which behaves much like a vowel in intrinsic velic positioning) is usually overlapped with, and visually indistinguishable from, the large movement for the nasal consonant (Figure 9a). That the vocalic portion is associated with a small lowering gesture underlyingly is supported not only by the normal rate data but also by corresponding minimal contrasts containing no nasal consonant, at the normal and rapid speaking rates (Figure 9b). Thus, the evidence supports the view that an increase in speaking rate is associated with greater overlap of gestures for adjacent (and near-adjacent) segments. Furthermore, cross-speaker comparisons showed a relation between speaking rate and extent of overlap.

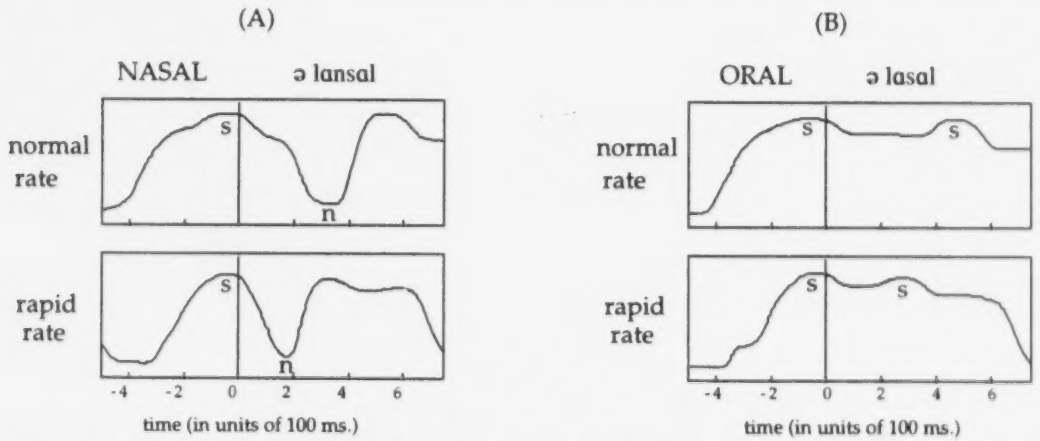


Figure 9. **(A) Sample velum movements for tokens of *It's a lansal again,* showing two-stage lowering between the /s/ of *It's* and the /n/ of *lansal* at a normal speaking rate (top panel) and single-stage lowering at a rapid rate (bottom panel). The vertical line in the middle of each panel marks the release of the /s/ of *It's,* as determined from the acoustic waveform. (B) Sample velic movements for tokens of *It's a lasal again,* showing velic lowering for the vocalic sequence following the /s/ of *It's* at both normal and rapid speaking rates. The velum rises again for the /s/ of /ləsəl/ at both rates. The vertical line in the middle of each panel marks the release of the /s/ of *It's,* as determined from the acoustic waveform. (The same speaker is represented in both A and B).**

That is, the speaker with the fastest self-selected "normal" rate in Bell-Berti and Krakow (1991a) produced most utterances with single-stage velic lowering (overlapping vocalic and nasal consonantal gestures), whereas the speaker with the slowest "normal" rate produced many more utterances in which two or more stages of lowering were evident.

Other studies on the effects of speaking rate show a reduction in the magnitude of velic movements at faster rates (e.g., Kent et al., 1974; Kuehn, 1976). Both types of observation (i.e., a decrease in multi-stage gestures and a reduction in positional extremes) are predicted by the coproduction model, as both are outcomes of increased temporal overlap of gestures, although no single study has looked for both in the same data. A reanalysis of the sequences with nasal consonants examined in Bell-Berti and Krakow (1991a), and listed in Table 3, supports the claim that the two types of observation are related effects of increasing gestural overlap. The position of the velum at the release of the /s/ of *It's* or *It's say* was measured for each sequence containing a nasal consonant in that study (Figure 10). The results show that the velic peak was consistently higher in the normal than in the fast rate productions, indicating that there was less overlap between the gesture for the nasal consonant (/n/) and that for the obstruent (/s/) at the normal than at the fast rate. In addition, the velic peak was higher when more vocalic segments intervened between the /s/ and the /n/, as shown in Figure 10. That is, increasing the duration of the vocalic string by slowing the rate or by adding segments, results in a larger separation between the extreme (opposite) positional requirements of the /s/ and of the /n/. The same manipulations have been shown by Bell-Berti and Krakow (1991a) to increase the likelihood that separate vocalic and nasal consonantal lowering movements will be seen (Figure 11).

Table 3. *Stimuli for rate experiment.*

	normal	fast
short	1. ɪts ansal	1. ɪts ansal
	2. ɪts lansal	2. ɪts lansal
	3. ɪts ə ansal	3. ɪts ə ansal
	4. ɪts ə lansal	4. ɪts ə lansal
	5. ɪts seɪ ansal	5. ɪts seɪ ansal
long	6. ɪts seɪ lansal	6. ɪts seɪ lansal
	long	short

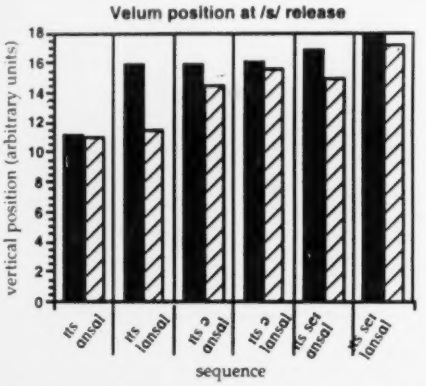


Figure 10. **Mean velum position at the release of the /s/ of *It's* or *It's say* obtained from 12 tokens of each of the sequences shown along the x-axis at each of two speaking rates. Normal rate sequences are represented with solid bars, rapid rate sequences, with striped bars. (The same speaker is represented here as in Figures 9 (A) and (B).**

The effects of rate manipulations on velic positional extremes have also been addressed in studies by Kent et al. (1974) and Kuehn (1976). Both of these studies suggest that subjects may vary in the manner in which they implement a rate change and lead to the conclusion that speaker-related differences are so robust and interesting that additional larger-scale studies of this issue are called for. For example, Kent et al. reported that both subjects in their study reduced the time taken to produce the sentence *Soon the snow began to melt* by about half when they switched from a moderate, conversational rate to a maximum, rapid rate (Figure 12). Although both also maintained the relative timing of velic peaks and valleys, the rate change was accomplished in different ways by the two speakers: One speaker (S1) increased the velocity of his faster utterances considerably, and showed a minimal change in movement extent. The other speaker (S2) showed little change in the velocity of his movements, but a considerable reduction in the movement extremes. Since the time taken to produce a sentence may be reduced by an increase in the velocity of movement, a decrease in movement extent, or by some combination of these changes, speakers may vary in their strategy. Similar differences among subjects are reported in studies of the effects of rate changes on lip and tongue movements (see, for example, Kuehn & Moll, 1976). Perhaps some speakers (or some speakers in particular situations) feel more constrained to maintain or approximate positional extremes than others; such speakers may be concerned with the effects of the rate change on intelligibility.

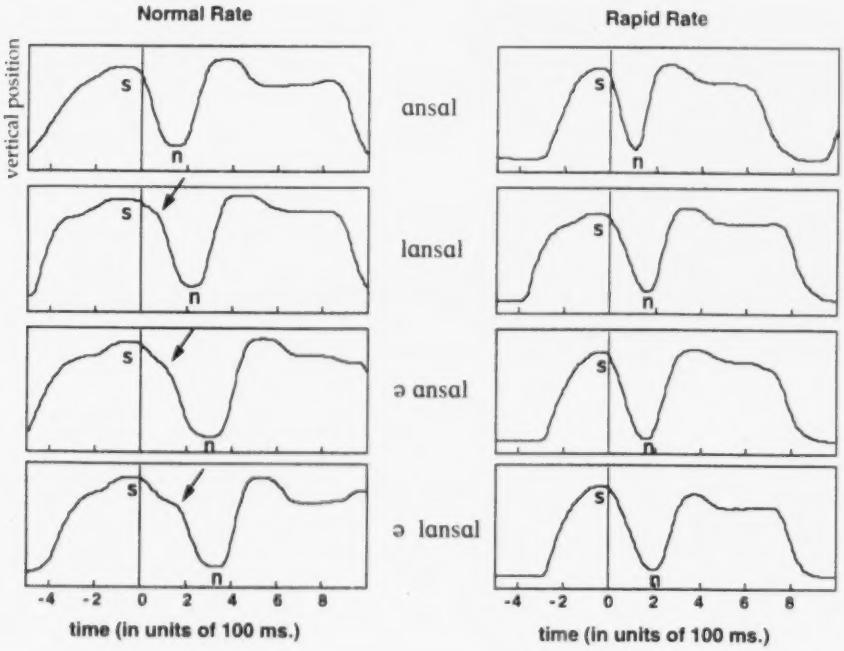


Figure 11. **Sample velum movements for sequences produced at normal (left panels) and rapid rates (right panels). The number of segments intervening between the /s/ of *It's* or *It's say* and the following /n/ increases from the top to the bottom panels of the figure. The arrows mark the emergence of a second stage of velic lowering.**

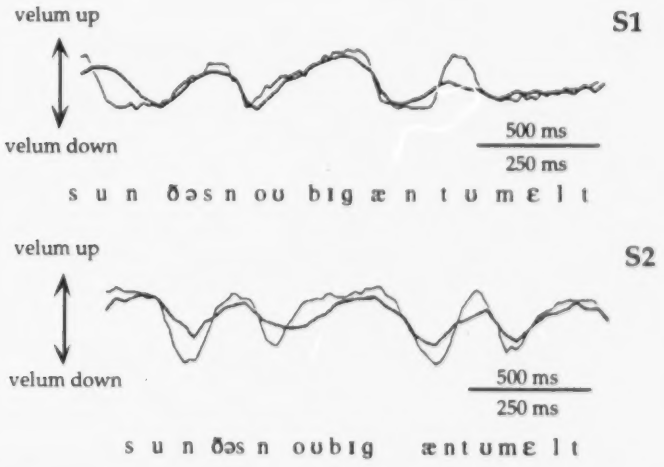


Figure 12. **Vertical velum movement patterns obtained from a cinefluorographic study as two subjects produced *Soon the snow began to melt* at moderate (thin lines) and rapid (thick lines) speaking rate . The rapid movements, which took about half the time of the normal rate movements, are time-normalized and overlaid on the normal rate movements. Two time scales are provided. Adapted, with permission, from Kent et al. (1974, Figure 9, p. 17).**

Another speaker-to-speaker difference in how rate affects velic movements was reported by Kuehn (1976), who showed that a reduction of positional extremes at rapid rates might, for some speakers, be observed only at the peaks or the valleys. The two subjects in Kuehn's study showed reduced range and duration of velic movement in rapid utterances that required opposite extreme velic positions in succession (e.g., VCNV or VNCV sequences). One subject (S1) reduced the range of movement by producing less extreme high positions (leaving low positions unaffected) whereas the other subject (S2) reduced the range by producing less extreme low positions (leaving high positions unaffected) (Figure 13).

Kuehn's account of the difference between the two subjects suggests that it is important to bear in mind that different speakers might use different parts of the total range of velic movement available at their normal rates. That is, Kuehn had noted that the speaker (S1) who reduced peak velic height in the faster condition had, at the normal rate, raised the velum beyond the level necessary to obtain sufficient velopharyngeal closure; the other speaker (S2) raised the velum just high enough to achieve adequate closure of the port at his normal rate. Hence, Kuehn's point is that, when asked to speak more quickly, S1 had raising to spare (so to speak), whereas S2 did not. For S2, the high peak could not be reduced without undesired acoustic consequences.

The studies done to date on velic movements at different speaking rates indicate that speakers may vary in the extent to which their fast productions approximate the extreme positions of their normal rate productions, since a rate change may be achieved by increasing the velocity and/or decreasing the range of movement. And, when a movement reduction occurs, it may affect both extremes of the normal range of movement, or just one.

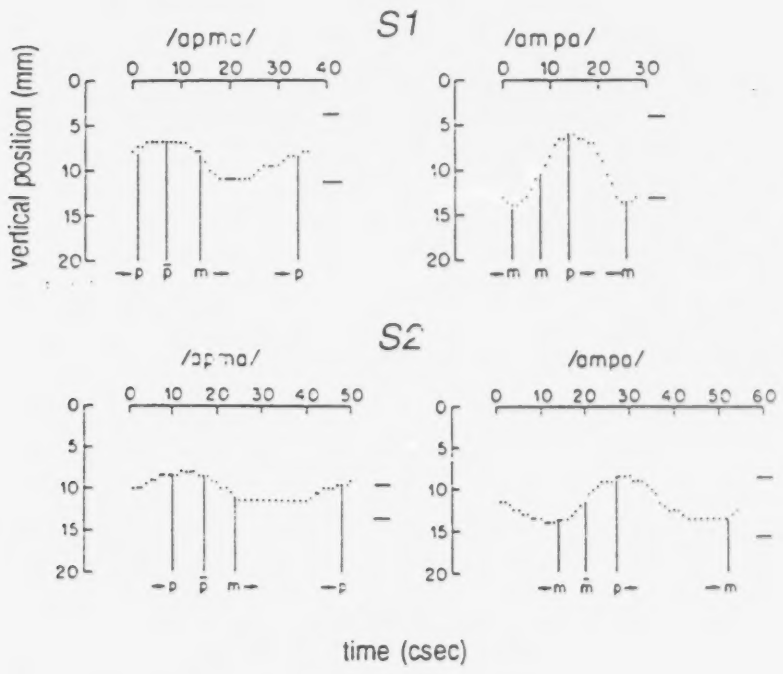


Figure 13. **Vertical velum movement patterns obtained from a cinefluorographic study for two utterances (/ɑpmɑ/ and /ɑmpɑ/) produced by two subjects at a rapid speaking rate. The horizontal lines at the right of each graph mark the vertical range of velic displacement during the same sequences produced at a normal rate. Adapted, with permission, from Kuehn (1976, Figure 7, p. 101).**

VI. DISCUSSION

Recent research has helped to elucidate the segmental organization of velic gestures, and to highlight the fact that differences between oral and nasal consonants are only one aspect of this organization; that is, there are also differences related to vowel height and to consonant place, manner, and voicing (see Bell-Berti, 1993, for a review). Furthermore, the research described in this chapter shows that, in addition to such segmental influences, there are influences of syllable organization, syllable position in a sentence, stress, and speaking rate.

Each of these nonsegmental influences was examined here in turn, but it was possible to see that velic movement patterns are affected simultaneously by a variety of segmental and nonsegmental influences. For example, the sequences examined in the section on stress (NV́C, NVC, CV́N, CVN) showed the combined effects of stress, nasal consonant position, and vowel height. The highest velic positions were found in /i/ following initial nasals—stressed /i/ for S1, but unstressed /i/ for S2. The lowest velic positions for both subjects were found in stressed /ɑ/ preceding final nasals. Despite subject-based differences in how stress affected velic height for /i/, the two subjects systematically distinguished the two vowels (/ɑ/ vs. /i/) in stressed vs. unstressed syllables in the contexts of initial vs. final nasal consonants.

Similarly, the declination data showed the overlapping effects of word stress and syllable position within a sentence. That is, although there was a general decline in the height of velic peaks over the course of a sentence, the decline was greater for stressed than unstressed syllables. In most cases, stressed syllables were characterized by higher velic peaks (for /t/) than unstressed syllables, but the greater decline in the stressed peaks meant that the contrast between stressed and unstressed syllables was reduced in later sentence positions. It can also be assumed that there are constraints on declination that are segment related. That is, reduction in the peak positions for /t/ over the course of a sentence must have limits: a certain minimum height is required to produce an obstruent.

Thus, identification of different nonsegmental influences on velic movements makes it possible to consider how the different segmental and nonsegmental influences combine to shape the movement patterns. However, a larger scope of investigation is still necessary because the relation between velic activity and the activities of other articulators must also be considered. As described in the sections on stress, declination, and speaking rate, it is important to note that certain patterns of velum movement have parallels in the patterns of other articulators. Such patterns may reflect general characteristics of the speech production mechanism. For example, the declination pattern observed in velic positions over the course of a sentence adds to a growing body of literature showing that a reduction in energy at later positions in a sentence occurs in laryngeal-respiratory behavior (from measures of F0, acoustic amplitude, and subglottal pressure) and in the activities of the articulators (from measures of the jaw, the velum, and vowel formant frequencies). Another example of a general pattern was found in the section on stress. That is, one subject increased the height of the velum for stressed /i/'s as compared to unstressed /i/'s, but decreased the height of the velum for stressed /ɑ/'s as compared to unstressed /ɑ/'s. This subject enhanced the intrinsic differences in velic height between the high and low vowels. The other subject, however, consistently lowered the velum in stressed syllables relative to the corresponding unstressed syllables, regardless of vowel height. It was noted that similar cross-speaker differences have been reported for tongue movements. Some speakers use higher positions for stressed than unstressed /i/, but lower positions for stressed than unstressed /ɑ/. These speakers are enhancing the intrinsic contrast between the tongue positions for high and low vowels. Other speakers use lower tongue positions for stressed vowels, regardless of height.

In addition to examining the combined effects of different segmental and nonsegmental influences on the velum, and to finding parallel patterns across articulators, it is also crucial to investigate those patterns which are specified by the coordination among articulators. For example, the coordination of velum lowering offset to the lip raising movement for a bilabial nasal consonant distinguished syllables with initial nasal consonants from those with final nasal consonants—across subjects, sequences, and tokens, regardless of whether the nasal consonants were at the middle or at the margins of a word. For syllable-initial bilabial nasals, the achievement of the low velum target and the high lower lip target coincided. For syllable-final bilabial nasals, however, the achievement of the low velum target was timed to the beginning of lower lip raising toward its target position.

This chapter, coupled with the chapter by Bell-Berti (1993), shows how much has been learned

about velic movement patterns in recent years. In contrast to earlier views of the velum as a functionally simple articulator, it is clear today that a variety of segmental and nonsegmental influences interact to shape velic movements. An understanding of these influences is critical to our understanding of the nature of speech motor organization and of the relation between speech motor organization and language sound structure.

REFERENCES

Beddor, P. S. (1982). *Phonological and phonetic effects of nasalization on vowel height*. Doctoral dissertation. University of Minnesota, Minneapolis. (Reproduced by the Indiana University Linguistics Club, 1983)

Bell-Berti, F. (1993). Understanding velic motor control: Studies of segmental context. In M. K. Huffman & R. A. Krakow (Eds.), *Phonetics and Phonology, Volume 5: Nasals, Nasalization, and the Velum* (pp. 63-85). New York: Academic Press.

Bell-Berti, F. (1980). Velopharyngeal function: A spatial-temporal model. In Lass, N. J. (Ed.), *Speech and Language: Advances in Basic Research and Practice* (Vol. 4). New York: Academic Press.

Bell-Berti, F., Baer, T., Harris, K., & Niimi, S. (1979). Coarticulatory effects of vowel quality on velar function. *Phonetica, 36*, 187-193.

Bell-Berti, F., & Harris, K. S. (1981). A temporal model of speech production. *Phonetica, 38*, 9-20.

Bell-Berti, F., & Hirose, H. (1975). Palatal activity in voicing distinctions: A simultaneous fiberoptic and electromyographic study. *Journal of Phonetics, 3*, 69-74.

Bell-Berti, F., & Krakow, R. A. (1991a). Anticipatory velar lowering: A coproduction account. *Journal of the Acoustical Society of America, 90*, 112-123.

Bell-Berti, F., & Krakow, R. A. (1991b). Velar height and sentence length: Declination? *Journal of the Acoustical Society of America, 89 (no. 4, part 2)*, 1916. (A).

Benguerel, A.-P., Hirose, H., Sawashima, M., & Ushijima, T. (1977). Velar coarticulation in French: A fiberoptic study. *Journal of Phonetics, 5*, 149-158.

Bladon, R. W. A., & Al-Bamerni, A. (1982). One-stage and two-stage temporal patterns of velar coarticulation. *Journal of the Acoustical Society of America, 72*, S104. (A)

Boyce, S. E. (1988). *The influence of phonological structure on articulatory organization in Turkish and in English: vowel harmony and coarticulation*. Unpublished doctoral dissertation. New Haven, CT: Yale University.

Breckinridge, J. (1977). *Declination as a phonological process*. Murray Hill, NJ: Bell System Technical Memorandum.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook , 3*, 219-252.

Browman, C. P., & Goldstein, L. M. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*, 140-155.

Browman, C. P., & Goldstein, L. M. (1991). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge, England: Cambridge University Press.

Chen, M. (1975). An areal study of nasalization in Chinese. In C. Ferguson, L. Hyman, & J. Ohala (Eds.), *Nasalfest*. Stanford: Stanford University Press.

Clumeck, H. (1976). Patterns of soft palate movements. *Journal of Phonetics, 4*, 337-351.

Cooper, W. E., & Sorenson, J. M. (1981). *Fundamental frequency in sentence production*. New York: Springer-Verlag.

Daniloff, R. G., & Moll, K. L. (1968). Coarticulation of lip rounding. *Journal of Speech and Hearing Research, 11*, 707-721.

Draper, M. H., Ladefoged, P., & Whitteridge, D. (1960). Expiratory pressures during speech. *British Medical Journal, 1*, 1837-1843.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*, 113-133.

Fowler, C. A. (1988). Periodic dwindling of acoustic and articulatory variables in speech production. *PAW Review, 3*, 10-13.

Fujimura, O. (1990). Methods and goals of speech production research. *Language and Speech, 33*, 195-258.

Fujimura, O., Miller, J. E., & Kiritani, S. (1975). An X-ray observation of movements of the velum and tongue. *Journal of the Acoustical Society of America, 58*, S40. (A)

Fujimura, O., Miller, J. E., & Kiritani, S. (1977). A computer-controlled X-ray microbeam study of articulatory characteristics of nasal consonants in English and Japanese. Paper presented at the 9th International Congress on Acoustics, Madrid, Spain.

Gay, T., Ushijima, T., Hirose, H., & Cooper, F. S. (1974). Effect of speaking rate on labial consonant-vowel articulation. *Journal of Phonetics, 2*, 47-63.

Gelfer, C. E. (1987). *A simultaneous physiological and acoustic study of fundamental frequency declination*. Unpublished doctoral dissertation. New York: CUNY.

Gelfer, C. E., Bell-Berti, F., & Harris, K. S. (1989). Determining the extent of coarticulation: Effects of experimental design. *Journal of the Acoustical Society of America, 86*, 2443-2445.

Hardcastle, W. J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication, 4*, 247-263.

Henderson, J. B. (1984). *Velopharyngeal function in oral and nasal vowels: A cross-language study*. Unpublished doctoral dissertation, University of Connecticut.

Horiguchi, S., & Bell-Berti, F. (1987). The velotrace: A device for monitoring velar position. *Cleft Palate Journal, 24*, 104-111.

Houde, R. A. (1967). *A study of tongue body movement during selected speech sounds*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Kahn, D. (1976). *Syllable-based generalizations in English phonology*. Bloomington, IN: Indiana University Linguistics Club.

Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics, 5*, 115-133.

Kent, R. D., & Netsell, R. (1971). The effects of stress contrasts in certain articulatory parameters. *Phonetica, 24*, 23 - 44.

Kent, R. D, Carney, P. J., & Severeid, L. R. (1974). Velar movement and timing: evaluation of a model of binary control. *Journal of Speech and Hearing Research, 17*, 470-488.

Kozhevnikov, V., & Chistovich, L. (1965). *Speech: Articulation and Perception*. Washington, DC: Joint Publications Research Service.

Krakow, R. A. (1987). Stress effects on the articulation and coarticulation of labial and velic gestures. Paper presented at the meeting of the American-Speech-Language-Hearing Association, New Orleans, LA.

Krakow, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. Unpublished doctoral dissertation, Yale University, New Haven, CT.

Krakow, R. A., Bell-Berti, F., & Wang, Q. (1991). Supralaryngeal patterns of declination: Labial and velar kinematics. *Journal of the Acoustical Society of America, 90 (no. 4, part 2)*, 2343. (A)

Krakow, R. A., & Huffman (1993). Instruments and techniques for investigating nasalization and velopharyngeal function in the laboratory: An introduction. In M. K. Huffman & R. A. Krakow (Eds.), *Phonetics and Phonology, Volume 5: Nasals, Nasalization, and the Velum* (pp. 3-59). New York: Academic Press.

Kuehn, D. P. (1976). A cineradiographic investigation of velar movement in two normals. *Cleft Palate Journal, 13*, 88-103.

Kuehn, D. P., & Moll, K. L. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics, 4*, 303-320.

Lightner, T. M. (1970). Why and how does nasalization take place? Papers in *Linguistics, 2*, 179-226.

Macchi, M. (1988). Labial articulation patterns associated with segmental features and syllable structure in English. *Phonetica, 45*, 109-121.

Maeda, S. (1976). *A characterization of American English intonation*. Unpublished doctoral dissertation. Cambridge, MA: MIT.

Matisoff, J. A. (1975). Rhinoglottophilia: The mysterious connection between nasality and glottality. In C. Ferguson, L. Hyman, & J. Ohala (Eds.), *Nasalfest: Papers from a Symposium on Nasals and Nasalization* (pp. 317-331). Stanford, CA: Stanford University.

Moll, K. L. (1965). A cinefluorographic study of velopharyngeal function during various activities. *Cleft Palate Journal, 2*, 112-122.

Moll, K. L., & Daniloff, R. G. (1971). Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America, 50*, 678-684.

Moll, K. L., & Shriner, T. H. (1967). Preliminary investigation of a new concept of velar activity during speech. *Cleft Palate Journal, 4*, 58-69.

Munhall, K., & Löfqvist, A. (1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics, 20*, 111-126.

Nittrouer, S., Munhall, K., Kelso, J. A. S., Tuller, B., & Harris, K. (1988). Patterns of interarticulator phasing and their relation to linguistic structure. *Journal of the Acoustical Society of America, 84*, 1653-1662.

Ohala, J. J. (1971). Monitoring soft palate movements in speech. *Project in Linguistic Analysis, 13*, J01-J015.

Ohala, J. J. (1975). Phonetic explanations for nasal sound patterns. In C. Ferguson, L. Hyman, & J. Ohala (Eds.), *Nasalfest: Papers from a Symposium on Nasals and Nasalization* (pp. 289-315). Stanford, CA: Stanford University.

Ohala, M. (1975). Nasals and nasalization in Hindi. In C. Ferguson, L. Hyman, & J. Ohala (Eds.), *Nasalfest: Papers from a Symposium on Nasals and Nasalization* (pp. 317-331). Stanford, CA: Stanford University.

Ohman, S. E. G. (1966). Coarticulation in VCV utterances. *Journal of the Acoustical Society of America, 39*, 151-168.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*, 333-382.

Schourup, A. (1973). A cross-language study of vowel nasalization. *Ohio State University Working Papers in Linguistics, 15*, 190-221.

Selkirk, E. O. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* Part II (pp. 337-383). Dordrecht, Holland: Foris Publications .

Stetson, R. H. (1951). *Motor phonetics: A study of speech movements in articulation* (2nd ed.). Amsterdam: North Holland.

Tatham, M. A. A. (1970). A speech production model for synthesis by rule. *Working Papers in Linguistics, 6*. (Computer and Information Services Research Center). Columbus, OH: Ohio State University.

Ushijima T., & Sawashima, M. (1972). Fiberscopic examination of velar movements during speech. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo, 6*, 25-38.

Vassiere, J. (1988). Prediction of articulatory movement of the velum from phonetic input. *Phonetica, 45*, 122-139.

Vatikiotis-Bateson, E., & Fowler, C. A. (1988). Kinematic analysis of articulatory declination. *Journal of the Acoustical Society of America, 84*, S128. (A)

Vayra, M., & Fowler, C. A. (1992). Declination of supralaryngeal gestures in spoken Italian. *Phonetica, 49*, 48-60.

FOOTNOTES

*Appears in *Phonetics and Phonology, Volume 5: Nasals, Nasalization, and the Velum* (Phonetics and Phonology Vol. V) (pp. 87-116). New York: Academic Press. (1993).

†Also Dept. of Speech-Language-Hearing, Temple University.

[1]The location of primary stress in these utterances was based on its fixed location in the single-syllable words of each set (i.e., each row in the table). The location of primary stress in the two-word items was matched to the one-word item in the same set. Thus, for example, because *homey* and *homely* have primary stress on the first syllable, so did *hoe me*, *home E*, and *home Lee* in this experiment. Since most of the items had primary stress on the first of the two syllables, a second set of utterances was constructed in Krakow (1989) to balance the position of stress. The results confirmed the findings of a bi-stable gestural organization for the syllable-initial vs. syllable-final nasals that is reported here, despite the presence of stress effects on other aspects of the movement.

[2]Lip contact data during /m/ were obtained with the use of an Electrolabiograph, a modified Electro-Glottograph, which supplied contact information on the lips, rather than the vocal cords. (For additional information, see Krakow, 1989).

[3]This velic declination study was limited to measures of peaks for obstruents. However, additional measures, including peak-to-valley displacements, ought to be examined in future studies. Such measures should shed light on whether there is centralization of both peaks and valleys in later sentence positions or whether there is some other pattern.

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 49-65

Articulatory Organization of Mandibular, Labial, and Velar Movements During Speech*

H. Betty Kollia,† Vincent L. Gracco, and Katherine S. Harris†

It has been shown that articulator movements during speech are adjusted along a number of spatiotemporal dimensions. For example, variations in the extent of lip, jaw, or tongue motion are associated with proportional changes in the respective articulators' peak velocity. Modifications in the timing of lip and jaw actions are apparently constrained, exhibiting relative timing covariation. Syllable prominence systematically affects some combination of the articulator motion parameters, i.e., extent, speed, and duration. The present investigation is an attempt to extend observations of the spatiotemporal properties of articulator movement to include the velum. Lip, jaw, and velar kinematics were recorded optoelectronically and simultaneously with the acoustic signal during productions of the utterance /mɑbnɑb/. The spatial and temporal relations between the lips, the jaw, and the velum were examined and compared across articulators. For movements associated with each syllable, the velum displayed scaling patterns qualitatively similar to those of the lips and jaw. Moreover, velocity-displacement relations were more robust for the lowering than for the raising movements of the velum. There was evidence of interarticulator coupling between the velum and the jaw, and between the velum and the upper lip, although this coupling was not as strong as that observed among the oral articulators. Articulator specific differences in velocity-displacement correlations and degree of interarticulator cohesion for the various movement phases may be related to a combination of aerodynamic and phonetic factors, such as the phonologically non-contrastive nature of nasalization in English.

INTRODUCTION

A number of investigations have shown that lip, jaw, and tongue movements during speech are modified systematically across segmental and suprasegmental variations. One of the more robust movement characteristics is the positive linear relationship of an articulator's peak displacement with its associated peak velocity. During speaking, reliable correlations have been observed in the kinematic parameters of various articulators, such as the jaw and the lip (Ohala et al., 1968; Kelso et al., 1985), the vocal folds (Munhall & Ostry, 1985), and the tongue dorsum (Parush, Ostry, & Munhall, 1983; Ostry, Keller, & Parush, 1983), to indicate that movement velocity is proportionally scaled to changes in movement extent. Further, this relationship between peak velocity and displacement is reliably observed regardless of movement direction (i.e., raising vs. lowering; Kelso et al., 1985; Munhall, Ostry & Parush 1985; Parush et al., 1983; Vatikiotis-Bateson, 1988; Vatikiotis-Bateson & Kelso, 1993). One of the aims of the present study was to examine the relationship between peak velocity and displacement in the velum.

No consistent direction-dependent trends have been noted in these relations in the literature. In a kinematic study of tongue dorsum movement, the peak velocity-displacement correlation for the raising movement toward closure was about the same as that for the lowering movement from the stop release (Parush et al., 1983). For one of the three subjects in that study, the raising movements were slightly faster than the lowering movements, and for another subject the raising movements were slightly larger than the lowering movements. In an analysis of speech kinematics of

This work is supported by NIH Grant DC-00121, and DC-00594 to Haskins Laboratories.

57

the jaw and the lower lip, Kelso et al. (1985) found that both opening and closing movements showed high correlations between peak velocity and displacement with a trend for the closing movements to display higher (steeper) velocity-displacement slopes than the opening movements. The same finding was also reported by Vatikiotis-Bateson and Kelso (1993), who concluded that, when viewed as a second order dynamical system, the peak velocity-displacement relationship can provide a sufficient spatial and temporal description of the overall movement behavior. A further aim of the present study was to explore direction-dependent trends in the relations between peak velocity and displacement in the velum, as well as in the jaw and the upper lip.

Stress assignment has been found to produce systematic changes in kinematic variables. Movements associated with stressed syllables are generally of greater displacement, higher velocity and longer duration than their unstressed counterparts (Gay 1968; Kelso et al., 1985; Kent & Netsell, 1971; MacNeilage, 1970; Ostry et al., 1983; Ostry & Cooke, 1987; Vatikiotis-Bateson, 1988; Vatikiotis-Bateson & Kelso, 1993). However, examination of the velocity-displacement relationships for various articulators reveals that the scaling of these two kinematic variables displays a consistent pattern in stressed versus unstressed syllables, that appears to reflect the varying movement durations. As the movement duration increases for the stressed element compared to the unstressed, there is a notable tendency for the velocity-displacement ratio to decrease (Munhall & Ostry, 1985; Ostry & Cooke, 1987). This trend has been observed for tongue dorsum movements (Ostry et al., 1983; Ostry & Cooke, 1987) as well as for jaw and lip movements (Edwards, Beckman, & Fletcher 1991; Kelso et al., 1985; Stone, 1981; Vayra & Fowler, 1992).

In contrast to the numerous studies of the jaw, lip, and tongue movement, relatively little research has focused on the movement characteristics of the velum. It is not known, for example, whether peak displacement and peak velocity scale in a similar manner for the velum as for the oral and laryngeal articulators, or whether stress prominence effects are reflected on velar raising movements. There is, however, a body of empirical evidence regarding certain determinants of velar position and velar movement. For example, Bell-Berti et al. (1979) found velar position to be influenced by vowel quality during adjacent nasal and oral consonants in utterances such as /fipmip/ or /fapmap/, and /fimpip/ or /fampap/. Velar position was lower in the environment of the low, open vowel /a/, than in that of the high, closed vowel /i/. Further, after a nasal consonant the velum was found to rise sooner and faster for a high vowel than for a low one. Similar findings have been reported by Moll (1962), Kent et al. (1974), and Clumeck (1976). Recently, Krakow (1993) reported stress and rate effects on movements of the velum, indicating that the velum may have an active role in the prosodic organization for speech. It was of interest, then, to cast a further glance on nonsegmental variables, such as stress prominence and utterance position, to determine whether their global effect on the kinematic parameters of the raising movements of the velum is comparable to that on lip and jaw kinematics.

In terms of interarticulator cohesion, Krakow (1989) examined the patterns of activity of the lower lip raising and of velum lowering, varying syllable position and word affiliation. She observed strong effects of syllable structure on the velum, but not on the lower lip. Velar movements showed consistent effects of syllable affiliation, and were generally amplified in syllable-final position. For syllable-initial /m/ the end of the lower lip raising movement preceded the end of the velar lowering movement, whereas for syllable-final nasals, the beginning of the lower lip raising movement preceded the end of the velar lowering movement. Thus, Krakow (1989) presented evidence of coordination between the movements of the lower lip and the lowering movements of the velum. McClean (1973) provided comparable observations of velar movements at junctural boundaries using cineradiography. To date there has been no detailed kinematic timing examination of the raising action of the velum with respect to the movements of other articulators. It is not clear whether and to what degree the velar raising movement is coupled to the movements of the jaw or the lips, when all articulators are actively involved in producing sound sequences. The present study aims to investigate these interarticulator timing relations.

Information about interarticulator cohesion has potential importance for understanding correlates of neuromotor organization for speech in theoretical questions that may be summarized as follows. If the previously observed (Gracco & Abbs, 1986; Gracco, 1988, 1994) consistency in articulator timing among the lips and jaw is found in movements of only that closely coupled group of articulators and did not include more distal articulators such as the velum, this might suggest an articulator organization reflecting local

constriction-producing events (i.e., oral, velar, laryngeal). Alternatively and more plausibly, if velar movements are found to demonstrate a degree of coupling with other articulators similar to the coupling observed between the lips and the jaw, that would suggest a control structure reflecting the functional demands of the system for speech production (Gracco, 1991, 1994). However, although a recent finding suggesting tightly coupled timing among the lips, jaw, and larynx during oral opening and oral closing provides some support for the latter position (Gracco & Löfqvist, in press), no comparable studies extending to the velum currently exist.

In the present experiment we investigated the functional linkages among the jaw, the lips and the velum, by examining the relative timing relations among raising and lowering actions of the velum, upper lip, and jaw. The hypothesis was that velar movement timing will be adjusted in conjunction with variations in lip and jaw movement timing. In sum, this study focused on the characteristics of velar movements within and across syllables in comparison to concomitant lip and jaw movements, and on the degree of temporal coupling among the velar and oral articulators.

Methods and Procedures

I. Subjects

The subjects were six females ranging in age from 25 to 50 years, with no known history of neurologic, hearing, or speech disorder. Except for BK, the first author, all subjects were native speakers of English and naive to the exact purposes of the experiment. AH, BK, CB, JP, and FE had had formal phonetic training, while LW had no formal phonetic training, but is a fluent speaker of a foreign language. The stimulus utterance obeys English phonotactic constraints. BK is a fluent speaker of English, and her productions were judged appropriate by phonetically trained native speakers of English.

II. Speech stimuli

The stimulus utterance used, /mɑbnɑb/, was selected because, for its production, the velum is expected to reach extremes of its range of movement (low for the nasal consonants and high for the stop consonant; Bell-Berti, 1973, 1976, Krakow, 1989). The velum begins rising from a lowered position for /m/ toward a high position for /b/, followed by rapid velar lowering for /n/, and a final rise for /b/. The vowel /ɑ/ was selected for maximal jaw and lip movements, as well as for maximal velar lowering since the coarticulatory effects of /ɑ/ on the nasal consonant result in a lower velar position than do the coarticulatory effects of other vowels. That is, the velum is at a lower position for /m/ in /mɑ/ than it is for /m/ in /mi/ (Bell-Berti et al., 1979).

The utterance was spoken at a self-selected conversational rate and subjects were instructed to place the primary stress of the utterance on the first syllable. The utterance was modeled for the subjects by the experimenter. As the structure of the utterance was possibly conducive to placement of almost equal stress on both syllables, any of the subjects' productions not conforming to the experimental specifications were discarded. The subjects were then alerted to the required stress pattern and were asked to repeat the token.

III. Instrumentation

III.a. The Velotrace

Velar movements were tracked using the Velotrace, a device developed for this purpose by Horiguchi and Bell-Berti (1987). The Velotrace consists of an internal lever connected to an external lever via a push-rod, which rides on a support rod, and is encased in a stainless steel tube. Part of the support rod rests on the floor of the nose, and part extends outside the nose. The tip of the curved internal lever rides on the nasal surface of the velum, moving with it; the movements of the internal lever are transmitted to the external lever through the push-rod. The external lever is nearly twice as long as the internal lever, and therefore the movements traced from the external lever are about twice as large as the actual velar movements that they reflect. The absolute magnitude of the Velotrace movements may vary across speakers; anatomical differences among subjects may result in differentially optimal positioning of the internal lever of the Velotrace, and, consequently, in different absolute displacements of the external lever. Speakers may also differ in the absolute extent of velar movement. The Velotrace is sensitive to even the very rapid movements of consecutive oral-nasal sequences.

III.b. The jaw splint

A custom fitted jaw splint was produced for each subject. A mandibular dental impression was taken and a plaster cast of the subject's mandibular dentition was made and used to mold the jaw splint. An acrylic resin casing of the lower teeth was made. The casing had two wide, stainless steel wires embedded in its outer edges. The wires were bent to exit the acrylic casing upward, at 45 degree angles, at the level of the cuspids, so as to not interfere with bilabial closure. One of the wires was then bent in a Z-shape, close to the chin to allow monitoring of the

jaw movement in the midsagittal plane. The other wire was bent horizontally and cut short, in order not to obscure monitoring of the upper lip movements. The jaw splint was kept in place with the help of a commercial dental adhesive.

The experimental set-up consisted of an adjustable dental chair, enclosed in an aluminum tubular beam framework, onto which a cephalostat was fixed. (For a more detailed description of this set-up see Kelso et al. 1984, pp. 814-815.) The Velotrace was then fastened to the stable parts of the cephalostat.

III.c. The LEDs

Infrared light emitting diodes (LEDs) were attached to the subject's lips, jaw splint and the external lever of the Velotrace in the midsagittal plane, and were tracked optoelectronically using a modified Selspot system.

IV. Procedure

The Velotrace was coated with 2% Xylocaine (Lidocaine HCl, a topical anesthetic) gel, and the subject's nasal cavity was sprayed with 4% Xylocaine spray. Even though no studies addressing the specific effect of topical anesthetic on movements of the velum exist, studies of the normal behavior of the larynx have revealed no discernible effect of topical anesthesia (Shipp, 1968 and Zemlin, 1969) on laryngeal activity. (See also Hardcastle, 1975 for the effects of topical anesthesia on the tongue.) The LEDs were placed on each subject at the following locations: the bridge of the nose (reference LED), the center of the vermilion border of the upper lip (upper lip LED), the tip of the external lever of the Velotrace (Velotrace LED), the external stable portion of the support rod of the Velotrace (Velotrace reference LED), and the point on the arm of the jaw splint closest to the subject's jaw (jaw LED).

The subjects repeated the stimulus utterance at a self-selected, comfortable speech rate following a cueing tone. Each subject produced a minimum of 50 tokens, as follows:

AH: 85,	BK: 69,	CB: 78,
FE: 78,	JP: 137,	LW: 50.

In the beginning and end of the experiment sustained /s/ and /m/ productions were obtained to view the maximum range of the velar displacement in order to ascertain optimal and functional positioning of the Velotrace. Each subject's data were analyzed separately.

V. Data Acquisition

The data were recorded on a multichannel instrumentation recorder at a speed of 3.75 inches per second for storage and subsequent analysis. A highly directional microphone (Sennheiser model MKH816T) was used for audio recording during the experiment. After the experiment, all movement signals were sampled from the tape using a laboratory computer with 12 bit resolution at a rate of 500 Hz per channel; the acoustic signal was low pass filtered at 4.8 kHz and sampled at 10 kHz. The horizontal and vertical positions of the two reference LEDs and the LEDs placed on the jaw splint and Velotrace were recorded along with the vertical position of the upper lip. For one subject (BK) the signal from the upper lip LED could not be used because the LED was intermittently obscured by one of the outer steel wires of the jaw splint. As a result, no upper lip kinematics were obtained for this subject.

The movement signals were smoothed using a 42 ms triangular window. The reference channels were subtracted from the appropriate kinematic channels to correct for any vertical head movement during the experiment. The resulting upper lip, jaw, and velum signals were then differentiated using a central difference algorithm to obtain the corresponding instantaneous velocities, which were subsequently smoothed using the same 42 ms triangular window.

VI. Data Analysis

The local maxima and minima of the movement and velocity signals for each token were marked automatically and individually inspected. The onset and offset of the movements were marked based on the zero-crossing values of the corresponding velocity signals. The coordinative timing of the velum, jaw, and lips were also examined. For these measures the time of occurrence for each articulator's peak velocity was identified for each movement phase and referenced to a common line-up point depending on the specific movement phase. For the first oral closing movement timing was referenced to the vowel onset in the first syllable; for the subsequent velar lowering/oral opening the occurrence of peak velocity was referenced to the peak jaw closing associated with the first /b/ in /mɑbnɑb/; for the final oral closing the oral articulators were referenced to the second jaw opening peak velocity associated with the onset of the vowel in the second syllable. Shown in Figure 1 are the acoustic signal for the utterance, along with the kinematic traces for the velum, the upper lip, the lower lip, and the jaw. The different movement phases are indicated by the shaded areas, and the vowel onset is marked.

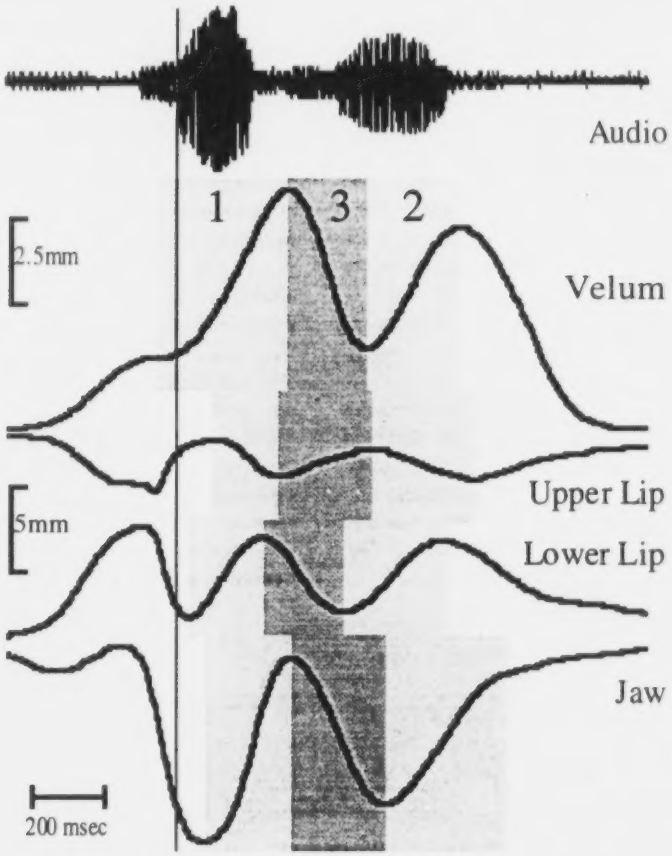


Figure 1. **Acoustic signal for /mɑbnɑb/, with the corresponding velar and oral movement kinematics. The velopharyngeal port is open during production of the nasal consonants /m/ and /n/, and is closed for the production of the bilabial stop /b/. For nasal sounds the velum is at a low position, while for oral sounds the velum is elevated. The first vowel onset is marked (thin vertical line), and shown are the corresponding averaged kinematic traces of the velum, the upper lip, the lower lip, and the jaw for tokens from subject AH. The shaded areas 1 and 2 indicate the first and second closing movements, and shaded area 3 indicates the intervening opening movements.**

RESULTS

In the present study, we examined the lip, jaw and velum movement characteristics associated with two contiguous syllables for a group of six subjects.

I. Movement characteristics: Velocity-Displacement relations

I.a. First closing

The velar movement for the first syllable was morphologically different from that of the lips and jaw, and different from that of the velar movement for the second syllable. As shown in Figure 1, the velar motion for the first syllable was a unidirectional movement (raising from a low to a high position, from /m/ to /b/), whereas the lip and jaw movements were composed of two distinct phases (an oral opening for the vowel, /m/ to /ɑ/, and an oral closing for the consonant, /ɑ/ to /b/). For the jaw and the lips, each opening and closing movement had a single associated peak velocity. However, this was not the case for some instances of the velar closing movement. Specifically, even though the velum was raised from /m/ to /b/, the velocity trace sometimes presented two distinct peaks. For all subjects some instances of two distinct velar raising movements were observed; one associated with the raising from the nasal to the vowel /ɑ/, and a subsequent raising for the consonant /b/. These multi-step movements were inconsistently present and generally occurred on the longer syllables, i.e., on the slower tokens. Similar multi-step movements were observed for velum lowering by Bell-Berti and Krakow (1991; see also Boyce et al., 1990).

Each velar raising step had its associated peak velocity. Since the first raising movement and its associated peak velocity correspond to the raising of the velum from the /m/ to the /ɑ/, and were not identifiable in all of the tokens analyzed, they were not used in the analyses presented here. Instead, the peak velocity associated with the raising of the velum from /ɑ/ to /b/ was used. This second velocity peak was comparable to the velocity peak of the jaw and lip movements for /b/ closure.

However, in these single-step velar raising movements the single velocity peak did not always occur at the same time in the velocity trace as the second velocity peak of the two-step jaw raising movement. Consequently, the temporal interval containing the velar peak velocity was longer than its oral counterpart, and, in fact, longer than the analogous interval for the second velar raising movement, which was invariably achieved in one step.

The first analysis focused on lip and jaw closing and velar raising for /b/. The velocity-displacement characteristics of the jaw closing and velar raising movements in the first stressed syllable were examined. On the left side of Figure 2 the jaw peak closing velocity and associated maximum closing displacement are presented for two of the six subjects. It can be seen that the peak velocity and displacement covary systematically. The velar peak velocity-displacement relations from the same context are presented on the right side of Figure 2 for the same two subjects. It is evident that, although the general relationship between velar raising movement velocity and displacement was comparable to that of the jaw, these variables were not as highly correlated for the velum as they were for the jaw. The same was true for the upper lip: movement velocity-displacement correlations were somewhat lower for the velum than for the upper lip, whose kinematic relations were comparable to the jaw's.

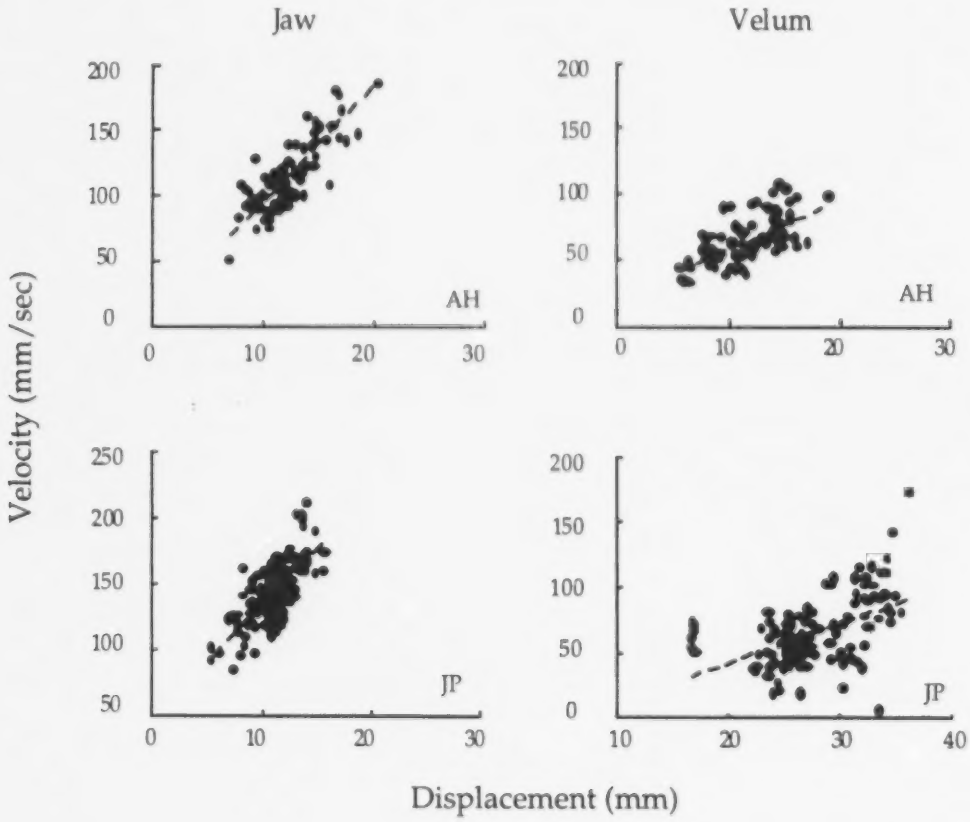


Figure 2. **Peak closing velocity - maximum displacement correlations for the 1st closing movement (/mɑb/), for the jaw and the velum for two of the six subjects. Peak closing velocity is plotted as a function of the displacement of the closing movement. Movement displacement is in mm and peak velocity in mm/sec.**

The increased variability in the velocity-displacement relations observed in the velum may be specific to the morphology of the first velar raising movement as discussed above. The trend for more variable velar velocity-displacement relations for the velum compared to the lip or jaw was seen in the data for all subjects. Presented in Table 1 are the velocity-displacement correlations for the upper lip, jaw, and velum for all subjects. The correlations for the upper lip and jaw ranged from r = .734 to r = .940. For velar raising, the correlations between peak velocity and peak displacement for the six subjects ranged from r = .138 to r = .773. All lip, jaw, and velar correlations were significant at the .01 level with the exception of the correlations in the velar movement for subject BK (see Table 1).

Table 1. *Velocity – displacement correlations for the 1st closing movement. (p<.01)*

SUBJECT	VELUM		JAW	UPPER LIP
AH	.625		.827	.734
BK	.138	NS	.881	---
CB	.350		.770	.800
FE	.773		.891	.746
JP	.494		.748	.940
LW	.305		.925	.753
mean	.485		.853	.815

I.b. Second closing

We next examined the correlation between the peak velocity and displacement for the second closing movement of the jaw, the velum, and the upper lip. The second syllable and, hence, closing movement differed from the first in terms of utterance position, phonetic context (/mɑb/ vs. /nɑb/), and stress prominence. Unlike the velum movements of the first syllable, the kinematic profile of the velum for the second syllable included both a lowering and a raising movement, thus rendering its morphology comparable to that of the jaw and lip. Moreover, in contrast with the first syllable, the velum raising movement in the second syllable was always achieved in a single step movement, even though this movement too involved raising of the velum for two separate "targets": from /m/ to /ɑ/, and from /ɑ/ to /b/. However, as a result of the single-step movement, only one evident velocity peak was associated with velar raising.

Figure 3 shows the peak closing jaw and velar velocity plotted as a function of the displacement of the closing movement for the /b/ in the syllable (/nɑb/) for two subjects. As shown in the figure and summarized for all subjects in Table 2, the velocity-displacement correlations for the velum were higher for the second closing than for the first closing. All correlations were significant ($p<.01$) except as noted.

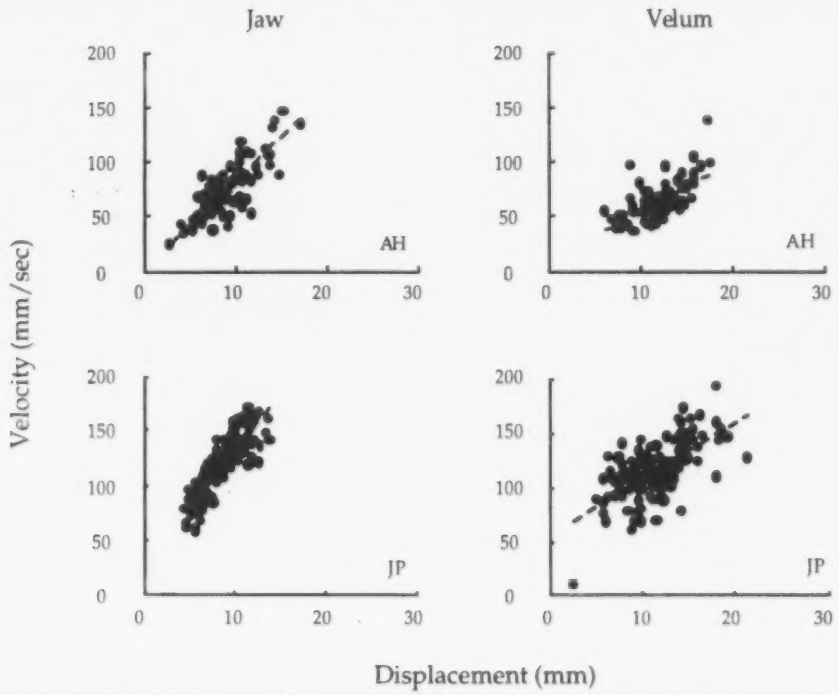


Figure 3. **Peak closing velocity - maximum displacement correlations for the 2nd closing movement (/nɑb/), for the jaw and the velum for two of the six subjects. Movement displacement is in mm and peak velocity in mm/sec.**

Table 2. *Velocity – displacement correlations for the 2nd closing movement. ($p<.01$ except where marked).*

SUBJECT	VELUM	JAW	UPPER LIP
AH	.627	.772	.835
BK	.308 $p<.02$	.916	---
CB	.634	.761	.753
FE	.885	.916	.601
JP	.621	.857	.900
LW	.875	.887	.341 $p<.02$
mean	.707	.865	.735

Table 3. *Velocity - displacement correlations for the opening movement /bnɑ/. ($p<.01$ for all correlation coefficients.)*

SUBJECT	VELUM	JAW	UPPER LIP
AH	.942	.884	.462
BK	.695	.906	.734
CB	.851	.762	.781
FE	.934	.914	.775
JP	.891	.866	.611
LW	.895	.945	.565
mean	.885	.890	.670

The upper lip and jaw velocity-displacement correlations for the second oral closing movement were comparable to those for the first closing movement except for LW's upper lip second closing movement. In this respect, and unlike the velum, the upper lip and the jaw showed similar velocity-displacement relations for the two closing movements, regardless of utterance position, phonetic context, or stress prominence.

I.c. Comparisons of the two closing movements.

Consistent with previous studies, in all articulators the closing movements for the stressed first syllable were larger compared to those of the second syllable. Figure 4 shows the average displacements for the first and second closing movements. For five of the six subjects, the movement displacements for the first syllable were significantly greater than those for the second syllable for the jaw and the velum ($p<.01$). For the upper lip, the results were less consistent. As shown in Figure 4 , the peak velocities also differed between the first and second closing movements for the jaw and the upper lip, and less consistently for the velum. The jaw and upper lip closing peak velocities were generally higher for the first syllable. For the velum, the trend was for the raising velocity to be equal or higher in the second syllable.

I.d. Opening movement

As mentioned above, for the first syllable (/mɑb/) the utterance began with the velum at a low position. For this reason we could examine only one velar lowering movement flanked by the first and second raising movements. Figure 5 presents plots of the peak velocity-peak displacement for the jaw and the velum lowering movements for two subjects. The jaw and velum lowering movements behave quite similarly as evidenced by the magnitude of their respective correlations. The peak opening velocity-displacement correlations for the jaw, the velum, and the upper lip for all subjects are presented in Table 3. All correlations were significant ($p<.01$).

To evaluate the possibility that the high peak velocity-displacement correlations observed in the velar lowering movement might reflect a mechanical effect related to the mass of the Velotrace rather than an active lowering action, we obtained the coefficients of variation for the lowering movement of the jaw and the velum. The reasoning was as follows: if the mass of the Velotrace were the major determinant of the lowering movement, then the characteristics of that movement (duration or velocity) would be extremely consistent, and, unlike the jaw lowering movement, would exhibit only a small degree of variability. We found, however, that the coefficients of variation (CV) for the lowering movements of the two articulators were relatively large. Moreover, the duration and velocity of the velar lowering movement was more variable than that of the jaw. The CV for the opening movement displacement ranged from 9.5 to 17.2 mm (mean 13.1) for the jaw, and 13.2 to 23.4 mm (mean 18.3) for the velum; peak velocity CV's ranged from 15.5 to 33.8 mm (mean 26.4) for the jaw, and 12.4 to 40.3 mm (mean 27.4) for the velum. It appears then, that the systematicity observed in the lowering movement characteristics of the velum is not an artifact of the presence of the transduction device.

I.e. Comparisons of movements in sequence

The velocity-displacement relations of the two closing movements were next compared to the velocity-displacement relations for the intervening opening action. Regression slopes were obtained from a simple least squares analysis relating the peak velocity and the displacement for each of the movement phases for the jaw, the velum, and the upper lip. In a simple sense, the slope measurements can be viewed as indicators of each articulator's movement frequency (see also Kelso et al., 1985). As such, it was possible to examine the global effects of variables such as utterance position, stress prominence, and movement

direction by comparing the slope changes within and across articulators. Table 4 presents the slopes obtained from the regression analysis. The velum showed the most consistent pattern across subjects, with the first closing movement having the lowest frequency, and the opening movement having the highest. The second closing movement was faster than the first, but slower than the opening.

For the jaw the opposite was true, the first closing was the fastest, and the opening movement was the slowest; the slope for the second closing tended to be lower than for the first closing, but this was not consistent across subjects. For the upper lip the results were not as systematic across subjects as they were for the jaw and the velum. For three of the five subjects the slopes of the closing movements were reduced from the first to the second syllable, while for the remaining two subjects (AH, JP) the slopes increased in the second syllable.

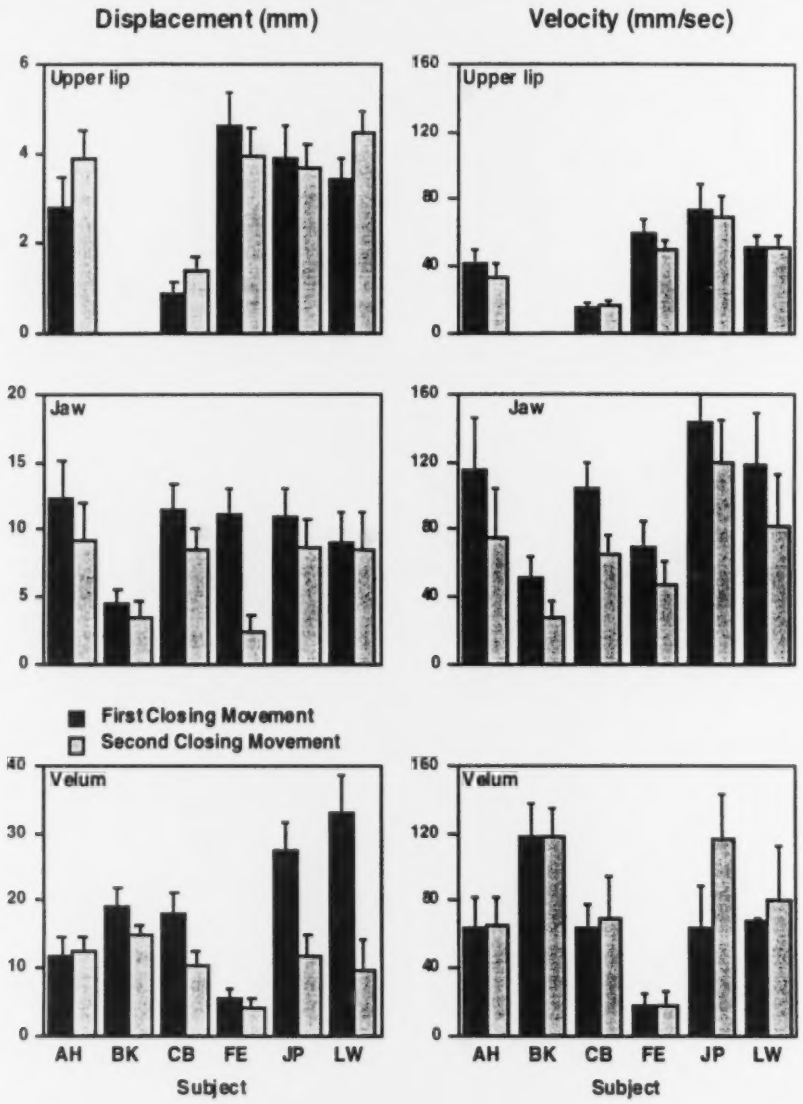


Figure 4. Mean /b/ closing displacements and velocities of articulator movements for the two syllables of /mɑbnɑb/ for the six subjects. The darker bars indicate the values for the first syllable, and the lighter bars the values for the second. The upper lip movement values are on the top panels, the jaw values in the middle, and the velum ones on the bottom panels. For the jaw, the first syllable is /mɑb/ and the second /nɑb/, while for the velum, the first syllable begins at the acoustic onset of the first vowel (/ɑb/), and the second begins at the offset of the stop closure (i.e., onset of the nasal: /bnɑb/). Displacements are in mm and velocities are in mm/sec.

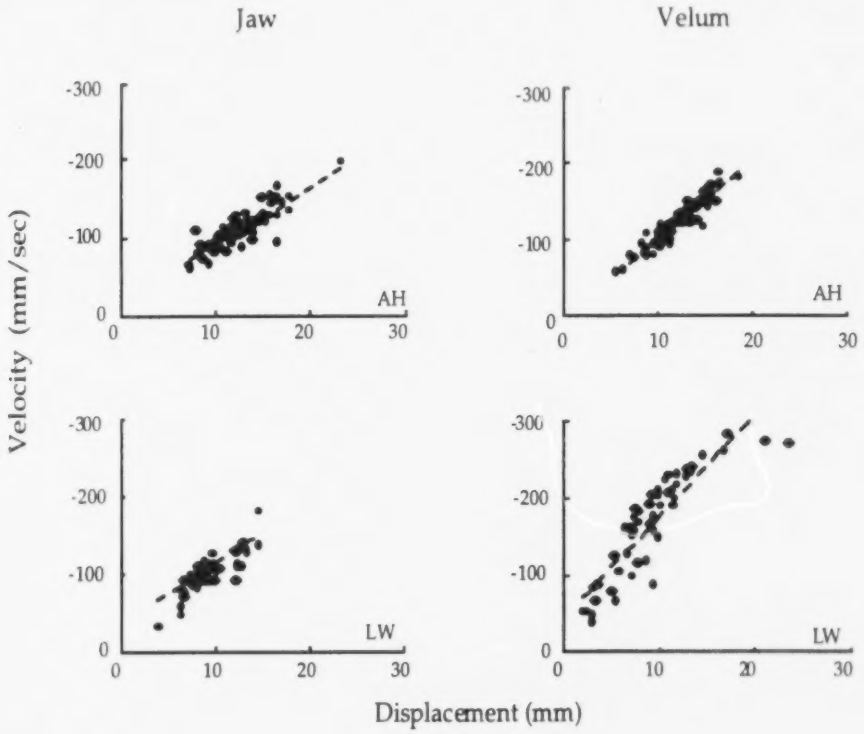


Figure 5. **Peak lowering velocity - maximum displacement correlations for the opening movement for the jaw (/bnɑ/) and the velum (/bn/) for two of the six subjects. Movement displacement is in mm and peak velocity in mm/sec.**

II. Interarticulator timing: Jaw-Velum-Upper Lip coordination

The second analysis focused on the relative timing relations among the velum, the jaw, and the upper lip for the first and second closing movements, as well as for the intervening opening movement.

II.a. Articulatory ordering

It is known that articulators cooperating in the same motor task demonstrate a high degree of temporal coupling to each other. As a first step to establishing whether there is any degree of temporal coupling between the velum and the jaw, we looked at the order with which the upper lip, the lower lip, the jaw, and the velum achieve peak position for closure. Generally the lips attained peak closing position first and the jaw and the velum followed. Beyond that, however, a more specific or stable ordering of the closure events was not observed. Regarding the order with which the same articulators arrived at peak position in the subsequent (/bnɑ/) opening movement, the only consistent observation was that the jaw reached peak opening position last, while the other articulators showed no systematic ordering pattern.

II.b. Relative timing: Comparisons between velar–oral and oral–oral articulator pairs

Previous studies have revealed a number of consistent kinematic relations among the movements of functionally related articulators. It has been shown, for example, that the lips and the jaw are coupled in their relative timing for oral closing (Gracco & Abbs, 1986; Gracco, 1988; McClean, Kroll, & Loftus, 1990; cf. also DeNil & Abbs, 1991). By analogy, we expected that, in the present study, the upper lip and jaw would demonstrate consistent relative timing of their kinematic landmarks and that the velum might assume a consistent relationship to them. Thus, in order to explore interarticulator cohesion, we examined the patterns of covariation exhibited in the kinematic behavior of the three articulators. The specific variables we examined were the times of attainment of peak velocity and peak position for the jaw, the upper lip, and the velum for the different movement phases. The degree of correlation between these variables was

considered an indicator of interarticulator cohesion.

II.b.1. First closing movement

For the first /b/ closing movement, the times of arrival at peak velocity and peak position were measured relative to the acoustic onset of first vowel in /mɑbnɑb/. As expected, a high degree of correlation between the timing of the jaw and the upper lip for the attainment of peak velocity was observed (see Table 5). The relation between the corresponding intervals for the velum and jaw showed greater variability than between the upper lip and jaw. One possibility for the apparently reduced cohesion among the velum and the jaw may relate to the particular velocity trace of the first raising movement, as was discussed in section I.a., earlier. A second possibility is that the relative timing for attainment of peak velocity for closure is not an important coordinating variable. Instead, a different temporal variable for the three articulators might demonstrate greater stability as an indicator of interarticulator coordination.

In order to examine this latter possibility, the relative timing of peak jaw, upper lip and velar positions were obtained for the first closing movement. Overall, the relative times of peak position rather than the times of peak velocity showed a higher correlation for the velum-jaw and the velum-upper lip closing movements. Interestingly, this was not the case for the upper lip-jaw relations. The relative timing results using peak position and peak velocity are presented in Table 5. In general, it seems that the relative timing of the closing events for the velum with the jaw and with the upper lip are related, but not as consistently as for the oral articulators with each other (i.e., the upper lip with the jaw).

Table 4. *Slopes of the regression equation of x on y, x=Displacement and y=Peak Velocity. Regressions were significant at the .01 level unless marked otherwise.*

	Closing 1			Opening 1			Closing 2		
SUBJECT	Velum	Jaw	Upper Lip	Velum	Jaw	Upper Lip	Velum	Jaw	Upper Lip
AH	3.7	8.8	9.7	10.4	6.4	6.5	4.4	7.9	10.5
BK	1.1 NS	10.2	---	13.6	8.3	---	4.4	6.7	---
CB	1.5	6.3	11.1	11.5	7.1	10.4	6.4	4.5	8.0
FE	5.9	10.4	9.0	12.2	9.5	14.7	5.5	10.7	6.1
JP	3.1	8.6	19.4	16.0	9.9	8.5	5.2	9.8	20.4
LW	1.3	12.4	11.4	13.2	6.2	9.4	6.0	9.1	5.3
mean	2.8	9.5	12.1	12.7	7.9	9.9	5.3	8.1	10.1

Table 5. *Times of peak position and peak velocity for the first /b/ closing movement: Correlations for Velum-Jaw and Upper Lip-Jaw. Correlations were significant at the .01 level. The /b/ closing time is measured from the time of the 1st vowel onset.*

	Time of peak position			Time of peak velocity		
SUBJECT	Velum - Jaw	Upper Lip - Jaw	Velum - Upper Lip	Velum - Jaw	Upper Lip - Jaw	Velum - Upper Lip
AH	.85	.62	.78	.32	.97	.30
BK	.78	---	---	.54	---	---
CB	.77	.87	.71	.13 NS	.94	.13 NS
FE	.59	.74	.69	-.01 NS	.91	.03 NS
JP	.36	.67	.51	.43	.91	.50
LW	.65	.89	.73	.37	.95	.41
mean	.695	.781	.694	.306	.941	.284

II.b.2. Second closing movement.

For the second closing movement, the referent used for measuring the times to peak velocity and peak position was the time of jaw peak velocity at /ɑ/ opening for /nɑb/. Examination of the closing movements for the second syllable showed a high correlation between the times of the attainment of peak closing velocity for the jaw and the upper lip (see Table 6). We found that the relative timing based on the attainment of peak position was somewhat better for the velum-upper lip pair than for the velum-jaw pair.

For subjects AH, BK, and FE the relative times of peak position for closure between velum-jaw and velum-upper lip were found to be more highly correlated than the relative times of their respective peak velocities. This was also true for subject CB in the timing relations between velum and upper lip. However, in the case of the velum-jaw relative timing the opposite was true for CB. Similarly, subject JP exhibited a higher degree of correlation in the relative timing between articulator peak velocities than between peak positions for velum-jaw and for velum-upper lip. For subject LW the timing of peak position for the velum did not correlate with either the time of peak position of the jaw, or with that of the upper lip. The relative timing results for peak position and peak velocity are presented in Table 6.

Overall, it seems that the relative timing of the velum for the second closing movement demonstrates some degree of temporal coupling with the oral articulators, particularly with the upper lip. For the second /b/ closing movement, the degree of coupling (as indicated by the correlations in the peak velocity times) among velar-oral articulators was not as high as that observed among the oral articulators (upper lip and jaw). However, the degree of coupling in the peak position times among velar-oral articulators was comparable to that observed among the oral articulators.

II.b.3. Opening movement

The relative timing between the velum and the jaw for the oral and velar opening movement in the second syllable was examined next. The intervals to peak position and peak velocity were measured from the peak jaw position at /b/ closure in /mɑb/. As noted earlier, velar lowering appears to be an actively controlled gesture. Therefore, the time of the peak velar lowering movement from /b/ to /n/ should covary with that of the jaw lowering from /b/ to /ɑ/ in the /nɑb/ context. Here, it was the timing of peak velocity across articulators (rather than of peak position as in the first syllable) that resulted in higher correlations. Moreover, the velum and jaw displayed the most consistent relations across the six subjects with all correlations being significant ($p<.01$). The correlation of the time of peak velocity between the velum and the jaw averaged r=.61 across subjects, ranging from r=.40 to r=.83. However, only two upper lip-velum correlations reached our significance criterion ($p<.01$) and another approximated it ($p<.02$). The times of peak velocity for the jaw and the upper lip were quite variable, with correlation coefficients ranging from r=–.12 to r=.67 across subjects (see Table 7). Similar findings have been reported previously (e.g., Gracco, 1988).

Table 6. *Times of peak position and peak velocity for the second /b/ closing movement (for /nɑb/). Correlations for Velum-Jaw, Upper Lip-Jaw and Velum-Upper Lip. /b/ closing time is with respect to the time of Jaw peak velocity for /ɑ/ opening in /nɑb/. Correlations are significant at the .01 level, unless otherwise indicated.*

SUBJECT	Time of peak position			Time of peak velocity		
	Velum - Jaw	Upper Lip - Jaw	Velum - Upper Lip	Velum - Jaw	Upper Lip - Jaw	Velum - Upper Lip
AH	.47	.40	.86	.35	.92	.30
BK	.63	---	---	.56	---	---
CB	.33	.54	.72	.46	.91	.44
FE	.29 $p<.05$	.61	.64	.28 $p<.05$	.95	.30
JP	.51	.70	.76	.73	.90	.82
LW	.13 NS	.67	.24 NS	.69	.73	.48
mean	.405	.592	.685	.531	.900	.505

Table 7. *Times of peak position and peak velocity for the /bnɑ/ opening movement: Correlations for Jaw-Velum and Jaw-Upper Lip. /bnɑ/ opening re: time of Jaw peak position for /b/ closing for /mɑb/. p<.01 except where marked otherwise.*

SUBJECT	Time of peak position			Time of peak velocity		
	Velum - Jaw	Upper Lip - Jaw	Velum - Upper Lip	Velum - Jaw	Upper Lip - Jaw	Velum - Upper Lip
AH	.74	.45	.53	.83	.67	.62
BK	.51	---	---	.59	---	---
CB	.42	.06 NS	.10 NS	.40	.13 NS	.05 NS
FE	.13 NS	-.14 NS	.12 NS	.42	-.12 NS	.27 $p<02$
JP	.53	.64	.56	.78	.37	.47
LW	.39	.40	.16 NS	.46	.36	.21 NS
mean	.475	.306	.310	.614	.306	.340

SUMMARY AND DISCUSSION

The purpose of this investigation was to examine the movement characteristics of the velum in a specific phonetic environment, across syllables and across concurrently active oral articulators, and to assess oral-velar interarticulator cohesion. Overall, the movement and relative timing characteristics of the velum were found to be similar to those of the jaw and the upper lip, although some consistent differences were seen in the velar raising motion. As has been previously observed for jaw and lip movement, the correlations between peak velocity and peak displacement were consistently high. For the velum, the peak velocity-displacement correlations for the two raising movements were statistically significant, but lower than the correlations observed in the oral articulators. In contrast to the raising movement, the lowering movement of the velum displayed higher velocity-displacement correlations than did the corresponding lowering movements of the lip and jaw. In terms of interarticulator cohesion, lip and jaw timing were found to covary as in previous studies of oral articulator coordination. In terms of remote articulators' intergestural cohesion, oral-velar closing movements showed a different pattern than oral-velar opening movements. The relative timing between the velar raising movements and the lip or jaw closing movements indicated less tight coupling than either oral-oral closing movements or oral-velar opening movements. The velar lowering movement, which displayed the most robust velocity-displacement correlations, also showed tight temporal coupling to the jaw lowering movement. The significance of these findings is discussed below.

I. Movement characteristics

I.a. Experimental considerations

Before addressing the results of the present study certain issues regarding the transduction technique and the methods used in the present study should be discussed. One of the findings of this investigation indicates that the movement and relative timing of the velum can be considered qualitatively similar to those of oral articulators such as the lips and jaw. In contrast to previous studies of lip and jaw movement and of relative timing, the movement characteristics of the velum were less consistent. One possibility accounting for the observed difference is that the presence of the device (the Velotrace) modified the velar movement patterns in a significant manner. Previous investigation of the characteristics of the device, however, suggests otherwise (Horiguchi & Bell-Berti, 1987). Horiguchi and Bell-Berti (1987) showed, for example, that the movements of the velum obtained with the Velotrace are qualitatively similar to comparable data obtained endoscopically from various subjects using the same speech material and that movement data obtained from simultaneous Velotrace and cineradiographic recordings are quantitatively similar. Specifically, they compared vertical velar movement measurements obtained cineradiographically to movement measurements of the tip of the internal level of the Velotrace. These movements were highly correlated (r=.90 and higher).

Another possibility is that the device, which only transduces velar motion in a plane, may not be capturing the complexity of the changes in velopharyngeal port area. While linear motion does not allow for accurate extrapolation to area measures due to the nonlinear relationship be-

tween planar movement and changes in area, it is improbable that measures of velar movement and area measures are unrelated. In fact, the displacement of the velum and fiberoptic measures of changes in port size have been shown to be highly correlated (Zimmermann, Dalston, Brown, Folkins, Linville, & Seaver, 1987). Specifically, Zimmermann et al. (1987) found correlations of r=.78 and r=.89 between photodetector output, indicating changes in port size, and displacement of the velum from the pharyngeal wall, measured cineradiographically. By analogy, a single point transduced at the midline of the lips or jaw does not allow for direct information on the position of all parts of either structure. Admittedly, reduced kinematic descriptions have certain limitations, yet results based on such simplified descriptions may, nonetheless, provide important information on articulator control principles.

I.b. Context effects

Despite the observed intra- and inter-subject variability, several generally consistent trends were evident in the data. Perhaps one of the most significant factors pertaining to velar and oral articulator differences is related to the morphology of the observed movements. For the oral articulators, the pattern of movement alternated as a function of the phonetic composition of the utterance. That is, the jaw started at a relatively high position for /m/, lowered for the first vowel, raised for the /b/ and /n/, lowered for the second vowel and raised for the final /b/. As detailed in the previous section, the velum started low for the /m/ and raised continuously through the first vowel, reached a peak for /b/, lowered for the /n/ and raised again for the second vowel and the final consonant. For the lip and jaw movements, the oral closing motion was always achieved in a single step, and was associated with a single phonetic segment. In contrast, the velar motion during the same interval progressed through two contiguous phonetic units.

The movement characteristics associated with velar raising were more variable than those observed for the lip and jaw, since the velar raising movement reflected a compound motion combining two velar gestures associated with two phonetic segments. Each segment was associated with a distinct velar position, with the position for the bilabial stop being higher than the position for the preceding vowel. This explanation is consistent with data and interpretations offered by previous investigators (Bell-Berti, 1976; Bell-Berti & Hirose, 1975; Bell-Berti, Baer, Harris, & Niimi, 1979; Boyce et al., 1990; Fritzell, 1969; Kent, Carney, & Severeid, 1974; Lubker, 1968; Krakow, 1989) that position of the velum is not specified simply in a binary manner (open vs. closed port), but in a continuous one, with the intermediate positions between maximally open (low) and maximally closed (high) being dependent on phonetic identity. Moreover, for the velar lowering movement associated with the single phonetic segment /n/, the velocity-displacement relations of the velum were comparable to those of the oral articulators. These findings underscore the contributions of contextual manipulations to investigations of articulatory characteristics.

Examination of effects such as utterance position and stress prominence on the articulators' closing movements indicated that they affected not only lip and jaw movements, as was expected, but also velar raising movements. As shown in Table 4, the slopes of the velocity-displacement correlations for the velar raising movements differed across syllables, apparently reflecting the degree to which the two steps of the raising movement are uncovered. In the first syllable, movement durations were somewhat longer, uncovering the two raising steps and resulting in weaker velocity-displacement correlations. In the second syllable, the raising movement was invariably achieved in one step, and the velocity-displacement correlations were higher than those seen in the first raising movement.

Certain direction-dependent trends were observed in the relations between peak velocity and displacement in the velum, as well as in the jaw and the upper lip. The movement frequency, as indexed by the slope of the velocity-displacement correlation, was higher in the closing movements than in the opening one for the jaw. The opposite was true for the velum: the movement frequency was higher in the opening movement than in the closing ones. It is possible that the difference observed between the jaw and velum trends is related to the morphology of the velar raising movements outlined earlier. As noted, the velocity-displacement correlations for velar lowering were generally higher than for the velar raising. While it seems possible that the velum motion was influenced by the mass of the lever of the Velotrace, subsequent analyses indicate that a mechanical interpretation of the findings is not tenable. The conclusion is, then, that velar lowering, like velar raising, is the result of controlled neuromuscular action, purposefully adjusted to the requirements of the phonetic environment and velar activity timing is regulated

along with other concomitant articulatory actions. This is consistent with interpretations offered by Bell-Berti and Krakow (1991).

II. Interarticulator timing: Upper Lip-Jaw, Velum-Jaw

The relative timing between upper lip and jaw movements for oral closure was highly consistent. This result is similar to findings of previous studies of lip and jaw temporal coordination (Gracco & Abbs, 1986; Gracco, 1994). The variations observed in vowel duration during the first and second syllables effected proportional changes in the timing of the two oral articulators. These changes, apparently aimed at maintaining a high degree of interarticulator cohesion, were observed only in the closing movement; the opening movement displayed no evidence of strong interarticulator coupling for lip and jaw timing (see also Gracco, 1988; 1994). Several findings regarding the coordination of the velum with the oral articulators are discussed below.

The relative timing relations of the velum with either of the oral articulators (lip or jaw) were somewhat weaker than those among the oral articulators. As outlined earlier, for the raising movements of the velum, the low velar-oral interarticulator correlations most likely reflect the differences in the movement pattern for the specific phonetic configuration. Given that the raising movement was a blending of two adjacent movements, the resultant velocity profile was not necessarily associated with a single segment. Instead, the velocity profile reflected the combined movement trajectory. Perhaps as a result, the time of peak velocity showed greater variability than the time of peak position, reducing the strength of the interarticulator correlations. Support for this explanation comes from the observation that, for the velar raising movements, higher correlations were found for the attainment of peak positions (rather than peak velocities) across the contributing articulators. In the case of peak positions, the intervals of interest were associated with the same phonetic target.

Overall, the velar-oral (velum-upper lip, as well as velum-jaw) correlations were somewhat lower than those found among the lip and jaw in this and previous studies. Comparisons of time of peak position vs. time of peak velocity across articulators demonstrated certain articulator- and parameter-specific differences. Velar-oral comparisons showed higher correlations for the timing of peak position than for the timing of the peak velocity. In contrast, for oral-oral (upper lip-jaw) comparisons higher correlations were found in the timing of peak velocities than in the timing of peak positions. Within the oral-oral closure framework, one explanation that may account for the relaxed oral-oral peak position timing relations compared to the peak velocity timing relations is that, for local constriction-producing events, like oral closing, the timing of lip and jaw movement toward closure (better indexed by the timing of peak velocity) is more critical than the timing of movement offset (or release, better indexed by the timing of peak position). In that case, the time of peak velocity is the measure more critically associated with the coordination of multiple articulators. For phonetic events such as velar-oral closure—given the morphological differences of the articulatory movements—the relative timing conditions may allow for a broader critical time frame of action constraint than for local events, thus providing an explanation of the observed patterns.

Consistent with this notion are the results from the oral-velar opening interactions for /bnɑ/; the stop closure is followed by a nasal consonant and low vowel requiring velum lowering and oral opening. It is reasonable to assume that the velar lowering for the nasal and the jaw lowering for the vowel should be tightly coupled in this context. For this portion of the second syllable the oral-velar interactions were indeed very strong. This was further evidenced in the higher correlations found in the times of peak lowering velocity compared to the times of peak lowering position. Recent findings from a study of the relative timing of the lips and larynx provide additional support for this notion (Gracco & Löfqvist, in press). Therefore, as interarticulator timing becomes more or less critical, the parameters that are indicative of coupling may also adjust: peak velocity relations may be more illustrative of critical or constrained timing relations while peak position timing relations may reflect a less constrained coordinative coupling.

In terms of contextual influences, a possible account for the somewhat relaxed peak position timing across velar-oral articulators is that vowel nasalization is not phonologically contrastive for any of the languages spoken by the subjects. Had vowel nasalization been controlled as a phonologically contrastive variable, the temporal interactions between velar and oral articulators would have been more highly constrained in order to maintain the contrast.

Examining these effects across syllables, it appears that oral-oral timing decreases from the first to the second closing movement, to the

intervening opening movement, with respect to peak position and peak velocity times. The oral-velar correlation pattern in peak position was nearly the reverse from that seen in peak velocity. Oral-velar peak velocity timing was less tightly constrained for the first than for the second closure, where the velum-lip and velum-jaw relative timing relations were more robust. Again, the reduced oral-velar peak velocity temporal cohesion for the first syllable is most likely related to the difference in the movement pattern for the first compared to the second syllable. For the opening movement, the oral-velar timing correlations were rather variable, but they were at least as high as the oral-oral timing correlations.

The general direction-dependent trends of interarticulator cohesion found in the data indicated that for closure events oral-oral cohesion was higher than oral-velar cohesion. For opening actions, however, oral-velar temporal coordination was better than oral-oral coherence.

In summary, the present investigation demonstrated that the kinematic characteristics of the velum are similar to those of the lips and the jaw. Velar velocity was scaled with velar displacement and the relative timing of velar actions showed adherence to the actions of the lips and the jaw. While there was a tendency for the timing covariation in the kinematic variables for the velar closing action to be less robust than that observed for the lips and jaw, the significance of this difference should be further explored in varying phonetic contexts.

REFERENCES

Bell-Berti, F. (1973). *The velopharyngeal mechanism: An electromyographic study*. Unpublished doctoral dissertation, City University of New York, New York.

Bell-Berti, F. (1976). An electromyographic study of velopharyngeal function. *Journal of Speech and Hearing Research, 19*, 225-240.

Bell-Berti, F., & Hirose, H. (1975). Palatal activity in voicing distinctions: A simultaneous fiberoptic and electromyographic study. *Journal of Phonetics, 3*, 69-74.

Bell-Berti, F. & Krakow, R. A. (1991). Anticipatory velar lowering: A coproduction account. *Journal of the Acoustical Society of America, 90*, 112-123.

Bell-Berti, F., Baer, T., Harris, K. S., & Niimi S. (1979). Coarticulatory effects of vowel quality on velar elevation. *Phonetica, 36*, 187-193.

Boyce, S. E., Krakow, R. A., Bell-Berti, F., & Gelfer, C. E. (1990). Converging sources of evidence for dissecting articulatory movements into core gestures. *Journal of Phonetics, 18*, 173-188.

Clumeck, H. (1976). Patterns of soft palate movements in six languages. *Journal of Phonetics, 4*, 337-351.

DeNil, L., & Abbs, J. (1991). Influence of speaking rate on upper lip, lower lip, and jaw peak velocity sequencing during bilabial closing movements. *Journal of the Acoustical Society of America, 89*, 845-849.

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America, 89*, 369-382.

Fritzell, B. (1969). The velopharyngeal muscles in speech: An electromyographic and cineradiographic study. *Acta Otolaryngologica, suppl.* 250.

Gay, T. J. (1968). Effect of speaking rate on diphthong formant movements. *Journal of the Acoustical Society of America, 44*, 1570-1573.

Gracco, V. L. (1991). Sensorimotor mechanisms in speech motor control. In H. Peters, W. Hultsijn, & C. W. Starkweather (Eds.) *Speech motor control and stuttering* (pp. 53-78). North Holland Elsevier.

Gracco, V. L. (1994). Some organizational characteristics of speech movement control. *Journal of Speech and Hearing Research, 37*, 4-27.

Gracco, V. L. (1988). Timing factors in the coordination of speech movements. *Journal of Neuroscience, 8*, 4628-4639.

Gracco, V. L.., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research, 65*, 156-166.

Gracco, V. L., & Löfqvist, A. (in press). Speech motor coordination and control: Evidence from lip, jaw, and laryngeal movements. *Journal of Neuroscience*.

Hardcastle, W. J. (1975). Some aspects of speech production under controlled conditions of oral anaesthesia and auditory masking. *Journal of Phonetics, 3*, 197-214.

Horiguchi, S., & Bell-Berti, F. (1987). The Velotrace: A device for monitoring velar position. *Cleft Palate Journal, 24*, 104-111.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C.A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 812-832.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America, 77*, 266-280.

Kent, R. D., & Netsell, R. (1971). Effects of stress contrasts on certain articulatory parameters. *Phonetica, 24*, 23-44.

Kent, R. D., Carney, P. J., & Severeid, L. R. (1974). Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research, 17*, 470-488.

Krakow, R. A. (1989). *The articulatory organization in syllables: A kinematic analysis of labial and velar gestures*. Unpublished doctoral dissertation, Yale University, New Haven, CT.

Krakow, R. A. (1993). Nonsegmental influences on the velum movement patterns: Syllables, sentences, stress, and speaking rate. In Huffman, M. K. & Krakow, R. A. (Eds.), *Nasals, nasalization, and the velum* (87-118). *(Phonetics & Phonology V)*. New York: Academic Press,

Lubker, J. (1968). An electromyographic-cineradiographic investigation of velar function during normal speech production. *Cleft Palate Journal, 5*, 1-18.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review, 77*, 182-196.

McClean, M. D., Kroll, R. M., & Loftus, S. N. (1990). Kinematic analysis of lip closure in stutterers' fluent speech. *Journal of Speech and Hearing Research, 33*, 755-760.

McClean, M. (1973). Forward coarticulation of velar movement at marked junctural boundaries. *Journal of Speech and Hearing Research, 16*, 286-296.

Moll, K. L. (1962). Velopharyngeal closure of vowels. *Journal of Speech and Hearing Research, 5*, 30-77.

Munhall, K. G., & Ostry, D. J. (1985). Ultrasonic measurement of laryngeal kinematics. In I. R. Titze & R. C. Scherer (Eds.), *Vocal*

fold physiology (pp. 145-162). Denver. Denver Center for Performing Arts.

Munhall, K. G., Ostry, D. J., & Parush, A. (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 457-474.

Ohala, J. J., Hiki, S., Hubler, S., & Harshman, R. (1968). Photoelectric methods of transducing lip and jaw movements in speech. *UCLA Working papers in Phonetics, 10*, 135-144.

Ostry, D. J. & Cooke, J. D. (1987). Kinematic patterns in speech and limb movements. In E. Keller and M. Gopnik (Eds.): *Motor and sensory processes of language* (pp. 223-235). Hillsdale, NJ: Lawrence Erlbaum Associates Inc..

Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 622-636.

Parush, A., Ostry, D. J., & Munhall, K. G. (1983). A kinematic study of lingual coarticulation in VCV sequences. *Journal of the Acoustical Society of America , 74*, 1115-1125.

Shipp, T. (1968). A technique for examination of laryngeal muscles during phonation. *Proceedings of the 1st International Congress of Electromyographic Kinesiology*, pp. 21-26, Montréal, Canada, August 1968.

Stone, M. (1981). Evidence for a rhythm pattern in speech production: Observations of jaw movement. *Journal of Phonetics, 9*, 109-120.

Vatikiotis-Bateson, E. (1988). *Linguistic structure and articulatory dynamics*. Bloomington, IN: Indiana University Linguistics Club.

Vatikiotis-Bateson, E. & Kelso, J. A. S. (1993). Rhythm type and articulatory dynamics in English, French and Japanese. *Journal of Phonetics, 21*, 231-265.

Vayra, M. & Fowler, C.A. (1992). Declination of supra laryngeal gestures in spoken Italian. *Phonetica, 49*, 48-60.

Zemlin, W. R. (1969). The effect of topical anesthesia on internal laryngeal behavior. *Acta Otolaryngologica, 68*, 176-196.

Zimmermann, G., Dalston, R. M., Brown, C., Folkins, J. W., Linville, R. N., & Seaver, E. J. (1987). Comparison of cineradiographic and photodetection techniques for assessing velopharyngeal function during speech. *Journal of Speech and Hearing Research, 30*, 564-569.

FOOTNOTE

[†]Also City University of New York Graduate Center.

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 67-79

An Acoustic and Electropalatographic Study of Lexical and Post-lexical Palatalization in American English*

Elizabeth C. Zsiga†

1 INTRODUCTION

In American English, alveolar obstruents (/t, d, s, z/) become palatoalveolars (/tʃ, dʒ, ʃ, ʒ/) before the (palatal) glide /j/. Palatalization is obligatory at the lexical level, as illustrated by pairs such as *habit / habitual*, *grade / gradual*, *confess / confession*, and *please / pleasure*. Palatalization also appears to apply, optionally, at the post-lexical level, as in the phrases *hit you*, *made you*, *press your point*, and *please yourself*. This paper argues, however, that lexical and post-lexical palatalization are two different processes, requiring two different representations.

Other investigators have noted that a similar process, palatalization of /s/ before /ʃ/, may be gradient when it applies across word boundaries (Catford, 1977; Shattuck-Huffnagel, Zue, & Bernstein, 1978; Zue & Shattuck-Huffnagel, 1980; see also Hulst & Nolan, in press, for British English). In the experiment reported here, acoustic and electropalatographic (EPG) data contrasting lexical and post-lexical palatalization of /s/ before /j/ were collected. The data show that palatalization of /s/ before /j/ is also gradient when it applies across word boundaries, while lexical palatalization is categorical. It is argued here that the articulatory patterns found in post-lexical palatalization suggest overlapping gestures, as in the theory of Articulatory Phonology (Browman & Goldstein, 1986, 1990, 1992). In a departure from Articulatory Phonology, however, it is further argued that lexical palatalization is best described in terms of features, and a mapping between features and gestures is suggested. Sections 2 and 3 discuss the methods and results of the experiment. Section 4 turns to the question of how the categorical and gradient rules should be represented.

This research was supported by NIH grant HD-01994 to Haskins Laboratories. I would like to thank Louis Goldstein, Draga Zec, and John McCarthy for invaluable input at all stages of this research; and Ken de Jong, Carol Fowler, Joaquin Romero, and the reviewers and editors of LabPhon IV for helpful comments on the manuscript.

2 Methods

2.1 Stimuli. Stimuli for this experiment contrasted underlying alveolars and palatoalveolars with both lexically-derived palatoalveolars and alveolar + /j/ sequences occurring across a word boundary (Table 1A). Data for /t/, /d/, /s/, and /z/ were collected, but only the data for /s/ are analyzed here. In the alveolar + /j/ sequences, /j/-initial pronouns and content words were contrasted, and the boundary between alveolar and glide was varied (phrase break vs. none). Each condition was represented by two different lexical items, divided into sets 1 and 2. In order to obtain information on the articulation of /j/, data from a second set of lexical items was also collected, in which the first consonant in the sequence was a labial (Table 1B). Within each set, the preceding vowel and the stress pattern remained constant across conditions.

Each lexical item was placed in a sentence. For presentation to the subjects, sentences were randomized within sets, over all consonants. For each subject five different randomizations of each set were created.

2.2 Data collection and analysis. Three native speakers of American English participated. Acoustic and EPG data were recorded simultaneously, using the Rion palatography system (Shibata, Ino, Yamashita, Hiki, Hiritani, &, Sawashima, 1978). In this system, palates are not custom made; rather, the best fit is chosen from six available sizes. The arrangement of electrodes on the palate is shown in Figure 1. One data frame (the 63 electrodes sampled in sequence) is recorded every 15.6 ms.

Table 1. *Stimuli contrasting underlying /s/, /ʃ/, and /j/ with lexically-derived /ʃ/ and /s#j/ sequences.*

	A. Fricatives:	SET 1	SET 2
1.	underlying /ʃ/	mesh on	fresh analysis
2.	underlying /s/ + /t/	confess to	press together
3.	underlying /s/ + /t/, phrase break	confess, to	impress, to get
4.	lexically-derived /ʃ/	confession	impression
5.	underlying /s/ + /i/	messy	dressy
6.	/s/ + you	confess you	press you
7.	/s/ + you, phrase break	confess, you	press, you
8.	/s/ + your	confess your	press your
9.	/s/ + your, phrase break	confess, your	press, your
10.	s/ + /j/-initial content word	confess unitedly	press uranium
11.	/s/ + /j/-initial content word, phr. break	confess, uniting	press, uranium
	B. Palatal glide:		
1.	/b/ + you	stab you	
2.	/b/ + you, phrase break	stab, you	
3.	/b/ + /j/-initial content word	stab Eugene	
4.	/b/ + /j/-initial content word, phrase break	stab, Eugene	

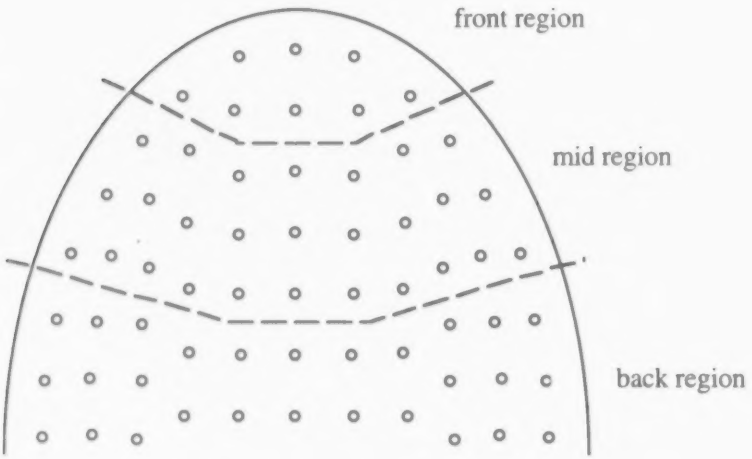


Figure 1. **Arrangement of electrodes on the palate, showing the division into front, mid, and back regions.**

In the acoustic analysis of the fricative tokens, the centroid of the fricative noise was computed for each EPG frame. (The centroid is the weighted average, based on amplitude, of all the frequencies present in the spectrum of fricative noise.) The times indicating the beginning and end of each EPG frame were marked in the acoustic signal, which was digitized at 20,000 samples/second. For each token, the spectrum of the fricative noise was computed from the first EPG frame showing full frication to the last, using a 12.8 ms Hamming window at 1 ms intervals. Centroid values over a range of 500 to 10,000 Hz were then computed for the window in the center of each EPG frame.

In the EPG analysis, patterns of palate contact for lexically-derived /ʃ/ and for the /s#j/ sequences were compared with patterns of contact in utterances containing underlying /s/, /ʃ/, and /j/. The series of EPG frames that made up each fricative or glide articulation was isolated on the basis of the articulatory patterns. For each control utterance (underlying /s/, /ʃ/, and /j/: 1 - 3 in Table 1A for the fricatives and 1 - 4 in Table 1B for /j/), an empirically-determined target pattern for the articulation was then located. Target was deemed to have been reached when the pattern of articulation remained stable over several frames (see Zsiga, 1993 for details). These target patterns formed the basis of templates, to which the other articulatory patterns were compared in terms of front, mid, and back contact.

3 Results

3.1 Acoustic results. Clear differences in the centroid values for the different fricatives were found. Figure 2 displays data for several sample utterances. These figures show the centroid values at each frame for five repetitions of underlying /s/, underlying /ʃ/, derived /ʃ/, and an /s#j/ sequence, aligned at the last frame of frication.

Figure 2A displays the centroid values for /s/ (in *press together*) and /ʃ/ (in *fresh*) for subject 1. The two fricatives are clearly distinct, and divide the figure into two regions, above 5200 Hz for /s/ and below 5200 Hz for /ʃ/. Lexically-derived /ʃ/ (Figure 2B) falls completely within the /ʃ/ region. The pattern is different, however, for /s#j/ (Figure 2C). In the phrase *press your*, the centroid values begin like /s/, but fall into the /ʃ/ region by the end.

Figure 2D-F shows the same contrasts for subject 2. As can be seen in Figure 2D, this subject shows even greater separation between /s/ (in *confess to*) and /ʃ/ (in *mesh*). Figure 2E shows that the /ʃ/ in *confession* is not different from the /ʃ/ in *mesh*. The /s#j/ tokens, however, show a lot of variation, and often a large change over time. While all are fully in the /ʃ/ range by the end of the fricative, at the beginning tokens are either /s/-like, or in between /s/ and /ʃ/.

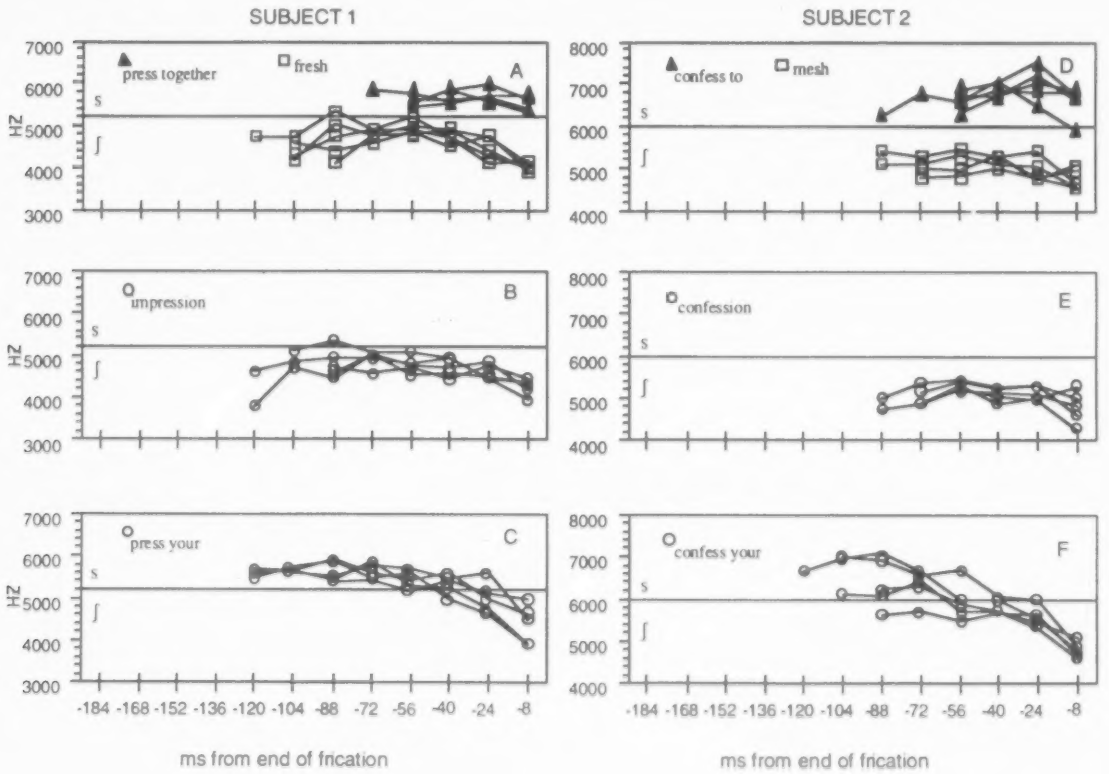


Figure 2. **Centroid values for several sample utterances.**

BEST COPY AVAILABLE

For statistical analysis, centroid values for all of the fricatives were compared at three points: the first frame of frication (onset), the last frame of frication (end) and the third to last frame of frication (-3 frames). The third from last, rather than the middle, frame was chosen for analysis because, as Figure 2 shows, the effect of a following articulation begins to be evident near the end of the fricative, but not necessarily halfway through. The hypotheses to be tested were (1) that lexically-derived /ʃ/ does not differ from underlying /ʃ/ and (2) that the /s#j/ sequences show partial palatalization, evidenced by falling centroid values over the course of the fricative.

These hypotheses were tested by comparing /s#j/ and lexically-derived /ʃ/ to underlying /s/ and /ʃ/ at the three measurement points. The two underlying fricatives are predicted to differ at each measurement point. It is predicted that lexically-derived /ʃ/ will not be distinct at any point from underlying /ʃ/, while the /s#j/ sequences will show a gradient change, with values not distinct from /s/ at the first frame, not distinct from /ʃ/ at the last frame, and possibly distinct from both /s/ and /ʃ/ at the third from last frame. Figure 3 shows the mean values for each subject for underlying /s/ (in *dressy* and *messy*), underlying /ʃ/ (*fresh*, *mesh*), derived /ʃ/ (*impression*, *confession*), and /s#j/ (averaged across all conditions with no phrase break). The /si/ rather than /s#t/ sequences were used in these comparisons to control for differences at the end of the fricative that might be due to the effect of a following stop rather than a more open articulation.

As is evident in Figure 3, the fricative types were different overall, and they differed in the way their values changed over time. In a repeated measures analysis of variance, with factors set, fricative type, and frame, each subject showed a highly significant effect of fricative type, and of the interaction between fricative type and frame. (No subject showed a significant 3-way interaction of fricativeXsetXframe.) As there was a significant effect of fricative for each subject at each frame, a Tukey test was used to analyze which fricatives differed from which at each of the three points. The predicted and actual contrasts among the different fricative types are shown in Table 2. In the table, conditions not significantly different at the .05 level are separated by an = sign, and where relevant (subject 3, end) are enclosed in identical bracketing.

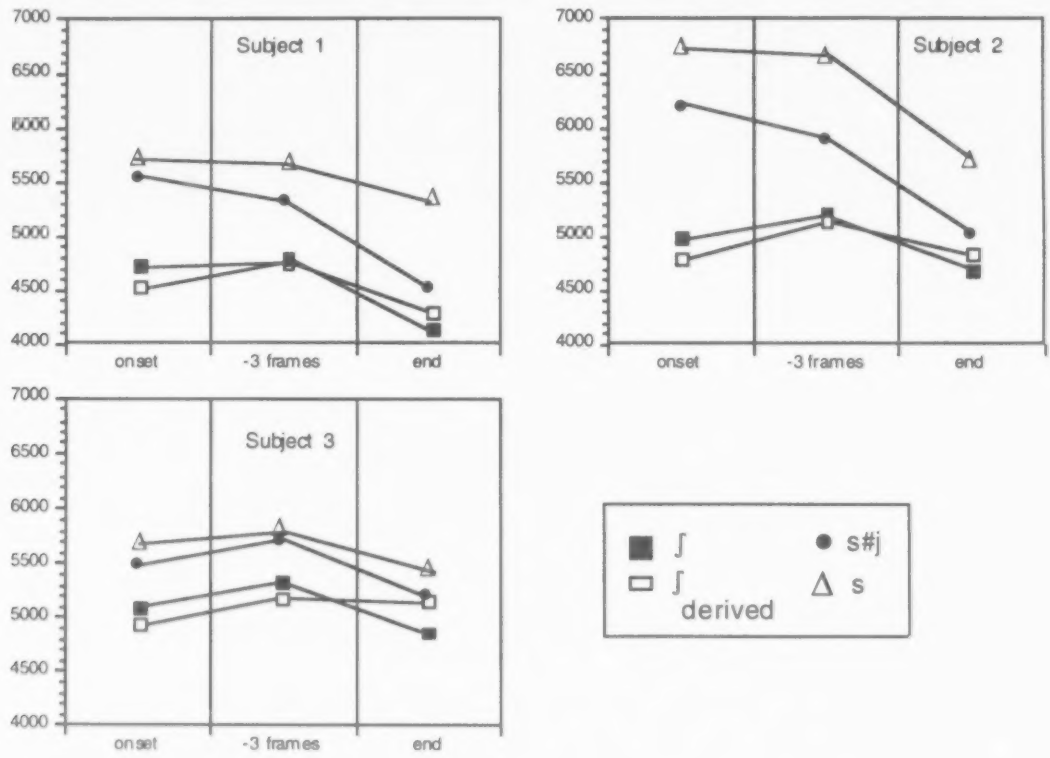


Figure 3. **Mean centroid values for four fricative types.**

Table 2. *Predicted and actual contrasts in centroid values at onset, -3 frames, and end for four fricative types.*

	onset	-3 frames	end
predicted:	s = s#j > ʃ-derived = ʃ	s > s#j > ʃ-derived = ʃ	s > s#j = ʃ-derived = ʃ
Subject 1:	s = s#j > ʃ-derived = ʃ	s > s#j > ʃ-derived = ʃ	s > s#j > ʃ-derived = ʃ
Subject 2:	s > s#j > ʃ-derived = ʃ	s > s#j > ʃ-derived = ʃ	s > s#j = ʃ-derived = ʃ
Subject 3:	s = s#j > ʃ-derived = ʃ	s = s#j > ʃ-derived = ʃ	(s= [s#j = ʃ-derived) = ʃ]

Underlying /s/ and underlying /ʃ/ were distinct at all three points for all three subjects. Underlying /ʃ/ and derived /ʃ/ were not distinct at any point for any of the subjects. The centroid values of these fricatives tended to be lower at onset and end (where amplitude was also lower) than at -3 frames.

In the /s#j/ sequences, subjects 1 and 2 fit the predicted pattern almost perfectly. For both subjects, although the centroid value for /s/ is lower at the end of the fricative than at -3 frames, the value for the /s#j/ sequences is lower still. For subject 1, there is substantial change over the course of the fricative in a /s#j/ sequence: at onset, /s#j/ is not significantly different from /s/; at -3 frames, /s#j/ falls in between /s/ and /ʃ/, and is significantly different from both; and at the end frame, /s#j/ is much closer to /ʃ/ than to /s/, although a significant difference remains between the sequence and underlying /ʃ/. Subject 2 also shows a large change over time. The /s#j/ sequence falls in between /s/ and /ʃ/ at both onset and -3 frames, although it is closer to /s/ at onset. At the end of the fricative, /s#j/ is not distinct from /ʃ/.

Results for subject 3 are less clear. Centroid values for underlying /s/ and /ʃ/ were more similar for this subject than for the other two, and overlapped to a greater extent. While the /s#j/ sequence shows lower values than /s/ throughout, it is not significantly different from underlying /s/ at any of the three points. At the end frame /s#j/ fell in between /s/ and /ʃ/, but is not significantly different from either. Note, however, that derived /ʃ/ is also not significantly different from /s/ at the end of the fricative. Derived /ʃ/ and /s#j/ have nearly identical values at this point, and even the difference between /s/ and underlying /ʃ/ was significant at the .05, but not the .01, level. This convergence of values at the end of the fricative makes the results for this subject difficult to interpret. He does show a tendency for /s#j/ centroid values to become /ʃ/-like at the very end of the fricative, but because the values for /s/ and /ʃ/ at this point are so close, /s#j/ and /s/ are not significantly different. It may be that placement of the palate interfered with articulation for this subject (see section 3.2).

Overall, these acoustic results show that lexical palatalization is categorical. There is no acoustic difference between underlying and derived /ʃ/. For /s#j/, palatalization is gradient, in two senses. First, /s#j/ shows substantial change over time, from /s/-like at the beginning to /ʃ/-like at the end. Second, the acoustics for the /s#j/ utterances may show centroid values intermediate between /s/ and /ʃ/ throughout the fricative. As could be seen in Figure 2, there is considerable token to token variation in the /s#j/ sequences. Some begin like /s/ and fall over time, some are /ʃ/-like throughout, others are in between the two. While lexical palatalization is categorical and obligatory, post-lexical palatalization is both gradient and variable in its application.

The next section turns to the patterns of articulation, and how these patterns are correlated with changes in the acoustics.

3.2 EPG results. Contact patterns at target for /s/, /ʃ/, and /j/ for each subject are shown in Figure 4. For /ʃ/, patterns are based on the steady state portion of the articulation in *mesh* and *fresh*, for /s/ in *confess to* and *press together* with no phrase break, and for /j/, in *stab you* both with and without phrase break. The number of times each electrode was activated at target over ten tokens is shown. Electrodes activated in 80% or more of the tokens are outlined. For the most part, the patterns shown here are qualitatively similar to those reported in earlier studies (e.g. Recasens, 1984, 1990; Hardcastle & Clark, 1981; Hardcastle, Gibbons & Nicolaidis, 1991). Some differences are discussed below.

For quantitative comparison, the palate was divided into three regions: front, comprising the first two rows; mid, comprising the middle three rows; and back, comprising the back three rows (see Figure 1). A one-way analysis of variance was performed on the target frames for each subject and region, with utterance as the independent variable and the number of electrodes activated within each region for each token as the

dependent. The utterance effect was highly significant ($p < .001$) for all subjects and regions, except for front contact for subject 3, where the effect just reached significance ($p = .036$). A Tukey test was then performed to determine which articulations were significantly different in each region. Very few differences were found due to the presence or absence of a phrase break, or to set. The /j/ in *eugene*, however, was found to show substantial coarticulation with the following affricate, and so is not used as a template (see Zsiga, 1993 for discussion). Differences between underlying /s/, underlying /ʃ/, and /j/ in *you* are summarized in Table 3. Differences not significant at the .05 level are separated by an = sign.

For subject 1, the target patterns are as expected for alveolars, palatoalveolars, and a palatal glide. The /s/ articulation shows more front contact and less back contact than any other: it is the only articulation showing contact in the frontmost row. /ʃ/ differs from /s/ in having less front contact, and from /j/ in having less back contact.

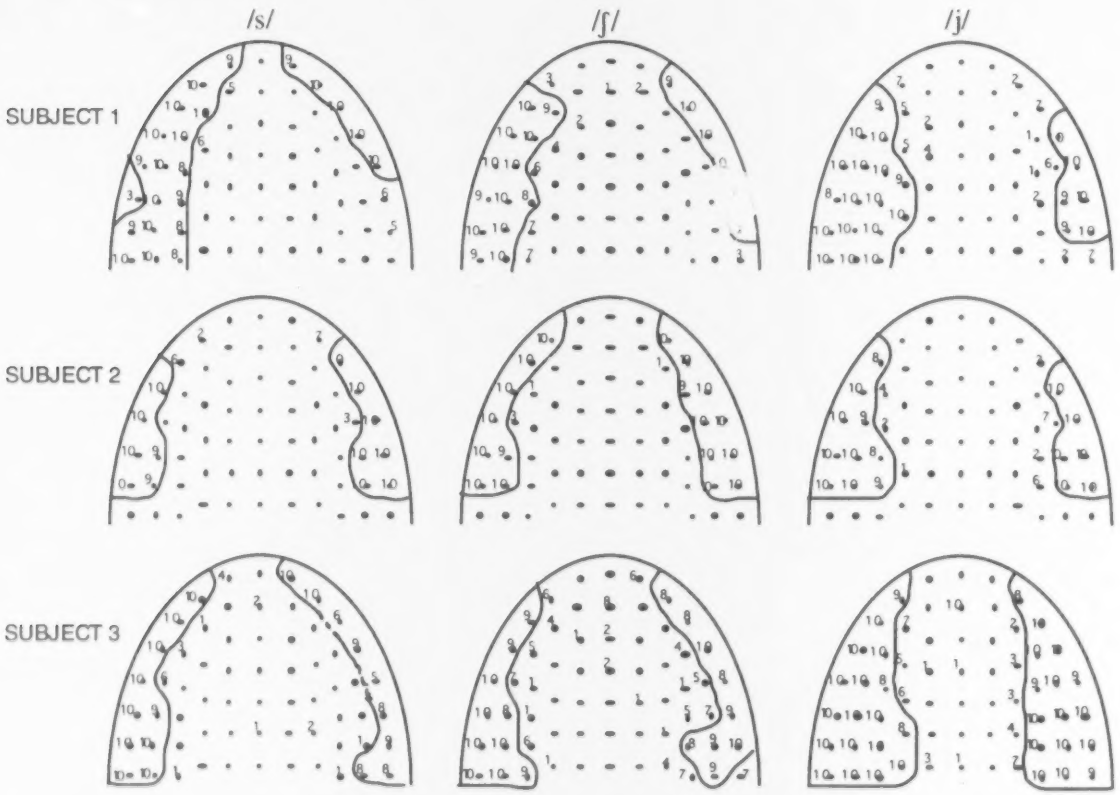


Figure 4. **Target patterns for underlying /s/, /ʃ/, and /j/ for each subject, based on a steady-state portion of the articulation.**

Table 3. *Differences in the number of electrodes activated at target in each region of the palate for each subject.*

	S1	S2	S3
front region	s > ʃ = j	ʃ > s = j	s = ʃ = j
mid region	s = ʃ = j	ʃ > s = j	s = ʃ = j
back region	s = ʃ < j	s = ʃ < j	s < ʃ < j

For subject 2, while /s/ and /ʃ/ show a significant difference in the amount of front contact, unexpectedly /ʃ/ shows the greater contact in this area. For this subject, the artificial palate was probably set slightly too far back, and failed to record contact at the frontmost edges of the subject's palate. (Recall that this subject showed the greatest acoustic difference between the two fricatives.) Due to the reversal in the amount of front contact for /s/ and /ʃ/ for this subject, and the fact that the two articulations do not differ in the amount of back contact, it is not clear that the patterns for /s/ and /ʃ/ can be reliably distinguished, nor is it clear how a /s#j/ sequence is predicted to differ from an underlying /ʃ/.

Subject 3 showed the least consistent patterns of articulation. As discussed above, his acoustic patterns were also problematic. Many electrodes are activated in fewer than half of the articulations. Several isolated electrodes, particularly in the center of the palate, appear to remain activated after release. (This can occur when the mouth is too dry.) For this subject, the articulations are clearly distinguished only in the back region.

Because of the difficulties in distinguishing the patterns in the control utterances for subjects 2 and 3, the rest of this discussion will focus on the articulatory data for subject 1. See Zsiga (1993) for a full discussion of all three subjects.

The patterns for underlying /s/, /ʃ/, and /j/ serve as templates for examining /s#j/ and derived /ʃ/. Figure 5 illustrates, for subject 1, how the patterns of activated electrodes in derived /ʃ/ (from *confession* and *impression*) and /s#j/ (from *press you* and *confess you*) change over time. Filled dots indicate those electrodes activated in at least 8 of 10 repetitions at the first frame of frication, the third to last frame, and the last frame. (Grayed electrodes were on 7/10 times.) Just as the acoustics do not change over time for derived /ʃ/, the pattern of palate contact remains stable throughout the whole fricative. The pattern for /s#j/, however, does not follow that for /ʃ/. At the onset of frication, there is very little contact at the back and center of the palate, but over the course of the fricative central and back contact fills in.

Figure 6 shows the pattern of electrodes for derived /ʃ/ and *s+you* at one point in time: -3 frames (the middle column in Figure 5). In each column of Figure 6, templates from the /s/, /ʃ/, and /j/ control utterances (corresponding to the outlined areas in Figure 4) are overlaid on the palate patterns. In the first column the template for underlying /ʃ/ is overlaid. This template corresponds almost exactly to the pattern for derived /ʃ/. For *s+you*, however, the /ʃ/ template is not a good fit: there is too much contact at the front and center of the palate. This poor fit illustrates that although the acoustics in the *s+you* sequence at this point in time may be /ʃ/-like, the articulation is not. Rather, as the set of template patterns in the second column shows, the pattern in the *s+you* sequence is what would be expected from an /s/ and /j/ being articulated at the same time. While the /s/ and /j/ templates do not fit the derived /ʃ/ articulation, they do account well for the front and central contact in the *s+you* sequence.

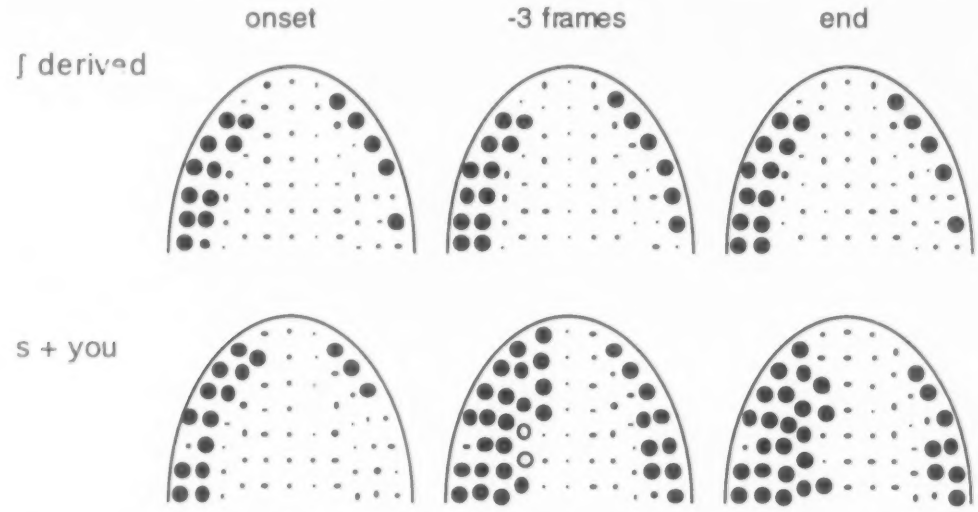


Figure 5. Change in contact patterns over time, subject 1. Electrodes shown were activated in at least 8 of ten repetitions.

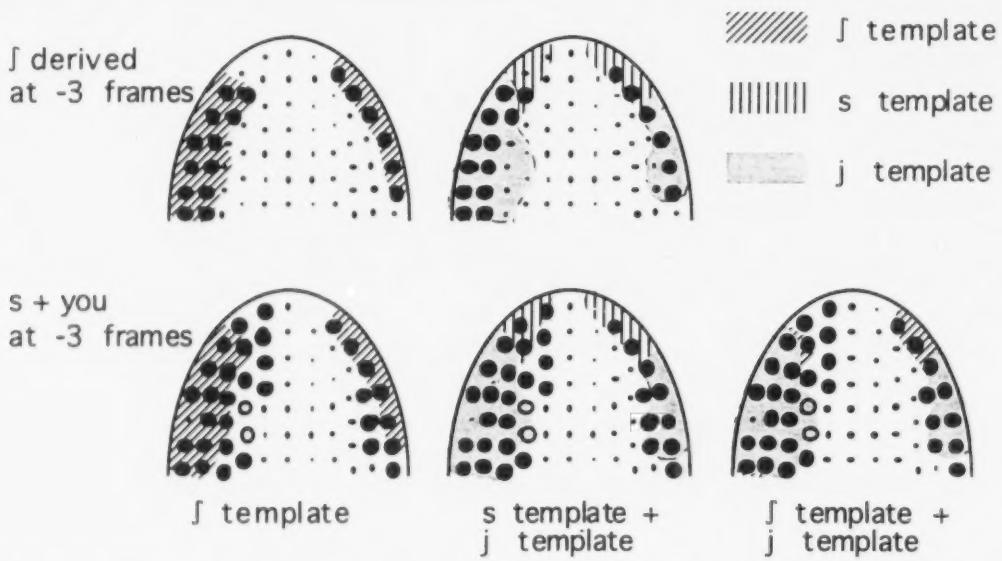


Figure 6. **Templates from underlying /s/, /ʃ/, and /j/ overlaid on the patterns for *s+you* and derived /ʃ/ at -3 frames.**

Although not all the *s+you* electrodes are covered by the templates, it is likely that the tongue, if it made contact at both the front and back regions at the same time, would cover the areas of those electrodes as well. In the third column, the /ʃ/ and /j/ templates are overlaid on the *s+you* pattern. It might have been the case that /s/ did undergo a categorical change to /ʃ/, and that the difference in the patterns seen in Figure 5 was due only to the fact that the following consonant is /j/ in one case but /n/ in the other. However, the combination of the /ʃ/ and /j/ templates could not account for the pattern of front contact seen for *s+you*. It is the pattern produced by the overlap of /s/ and /j/ that fits the *s+you* articulation at -3 frames most closely.

Finally, a significant correlation was found between the acoustic and articulatory measures. The centroid values at onset, -3 frames, and end were correlated with the total number of electrodes activated in the front and back regions at those points. For subject 1, the amount of back contact accounts for 28% of the variance in the centroid values, and front and back contact, taken together, account for 45% of the variance (both significant at $p < .01$). Back contact, not front contact, better determines the centroid value. As contact in the back region increases, the centroid value falls. These findings are consistent with the hypothesis that in the /s#j/ sequences, increased overlap of the /s/ and /j/ gestures, and therefore more back contact during the fricative, leads to the lower centroid values.

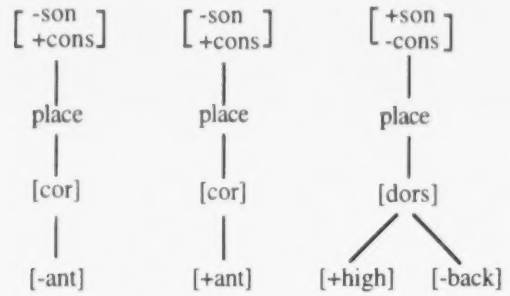
4 Discussion

This experiment has shown a clear difference between lexical and post-lexical palatalization in American English. Post-lexical palatalization is gradient and variable. Lexical palatalization, on the other hand, involves a categorical alternation between /s/ and /ʃ/: underlying and derived /ʃ/ were not found to differ either acoustically or articulatorily. (The data presented here is consistent with the view that the coronal fricative in *confession* in fact *is* an underlying /ʃ/. It will be assumed here, however, that the regular lexical alternations relating words such as *confess* and *confession* should be expressed in the grammar as phonological rules.) This section examines the question of how the two different kinds of palatalization should be represented. Representations using autosegmental features and articulatory gestures are compared. It will be argued that both representations are needed: phonological features best capture categorical alternations, articulatory gestures best capture gradient processes.

Consider first the featural representation. The featural representation of American English palatalization is bound up with the question of the best way to represent palatal consonants and glides. In the feature geometries argued for in

Sagey (1986) and McCarthy (1988), /s/ and /ʃ/ are represented with the feature [ant] as a dependent of the [coronal] node (1a, 1b). The glide /j/ is represented with the features [+high] and [-back] as dependents of the dorsal node (1c).

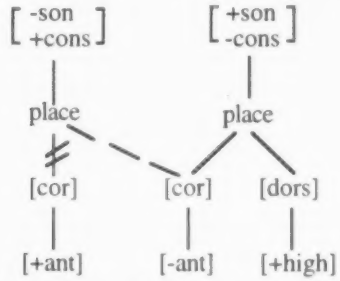
(1) a. /ʃ/ b. /s/ c. /j/



Borowsky (1986) formalizes /s#j/ palatalization as spreading of [+high] from the glide to the alveolar. In this formalization, a dorsal node to which the feature can attach must be interpolated, and then a special implementation rule must be invoked to interpret the resulting configuration as phonetically identical to that in (1a).

However, recent studies, both phonological (Clements, 1976, 1991; Hume, 1990, 1992; Broselow & Niyondagara, in press; Ní Chiosáin, 1991) and phonetic (Keating, 1988), have argued that /j/ should be analyzed as having a coronal component. (For an opposing point of view, see Recasens, 1990, this volume.) The representation of palatalization in (2) follows Keating in representing /j/ as a complex segment with both coronal and dorsal components. The [-ant] coronal component spreads from the glide to the alveolar, effecting a categorical change from /s/ to /ʃ/.

(2)



As will be argued below, however, this representation is not appropriate for gradient palatalization.

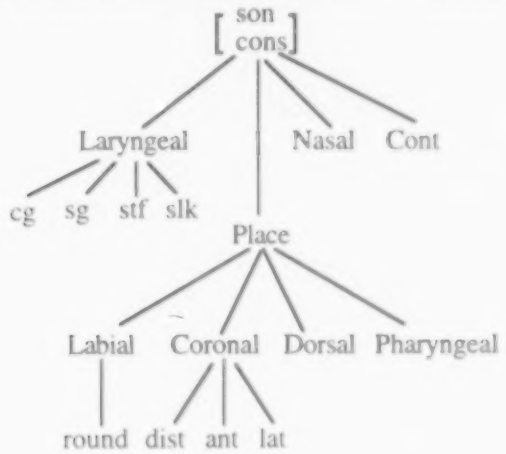
Consider instead the gestural representation. The articulatory evidence presented here (at least for the subject for whom clear results can be obtained) suggests that gradient palatalization can be represented in terms of gestural overlap (following Browman & Goldstein, 1986, 1990, 1992). The palatal constriction for /j/ overlaps in time with the alveolar constriction for /s/ when /s/ and /j/ are adjacent at a word boundary. The combination of front contact due to the /s/ gesture and increasing back contact due to the /j/ gesture results in centroid values that fall over the course of the fricative.

The gestural representation captures the gradience and variability seen in post-lexical palatalization. For the period of time before the /j/ gesture begins, the articulation and acoustics will be that of a simple /s/. As overlap with the /j/ increases, the articulatory and acoustic influence of this gesture also increase. Many instrumental studies have demonstrated overlap among speech gestures (e.g. Hardcastle, 1985; Hardcastle & Roach, 1977; Öhman, 1966; Perkell, 1969; and Marchal, 1988). Zsiga (1994) provides evidence for substantial overlap between consonant gestures at word boundaries in English. It may be that the pattern of overlap seen here is just that typical of the overlap between any two consonants at a word boundary. If so, then no post-lexical rule of palatalization is required. The effect of palatalization would simply be the acoustic consequence of the normal pattern of overlap.

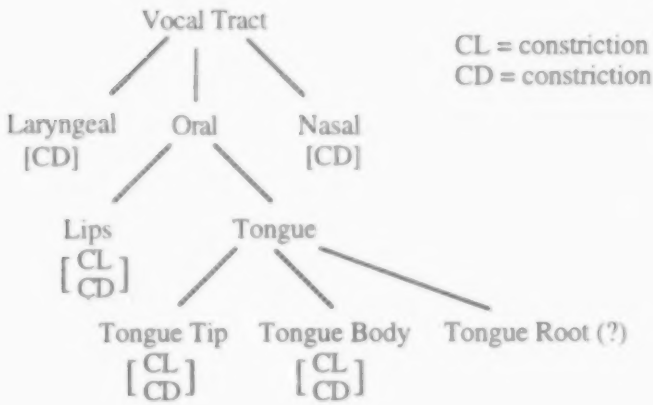
Both categorical and gradient palatalization can be seen as the imposition of the high tongue position for the glide onto the consonant. In categorical palatalization, the [-ant] coronal feature spreads from one root node to the next. In gradient palatalization, the /j/ gesture overlaps in time with the /s/ gesture.

The featural and gestural representations in fact correspond very closely. Compare the autosegmental representation in (3), argued for in McCarthy (1988), with (4), the "functional anatomy of the vocal tract" presented in Browman and Goldstein (1989). Both representations are based on articulators: the lips, the tongue tip, the tongue body, and possibly the tongue root. (See Browman & Goldstein, 1989; McCarthy, in press for discussion of the representation of pharyngeal articulations.) In both the gestural and the autosegmental representations, the nasal, laryngeal, and oral subsystems are separated. Browman and Goldstein (1989) have pointed out that the convergence on a single geometry from the direction of phonological patterning and from the direction of phonetic function provides strong support for the geometry's essential correctness.

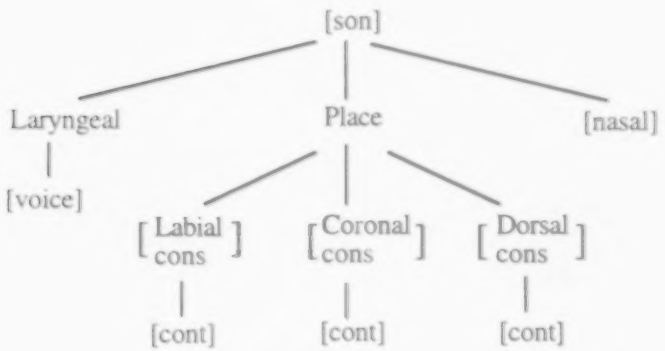
There is an even more striking similarity to the feature geometry proposed in Padgett (1991). Padgett argues, on the basis of patterns of assimilation, that [continuant] and [consonantal] should be specified for each articulator, in an "articulator group", as shown in (5). These two features then correspond directly to the constriction degree and stiffness specified for each gesture. (In Articulatory Phonology, stiffness encodes the difference between vowels, glides, and consonants.) While smaller differences between the two geometries remain to be resolved (see Zsiga, 1993), a straightforward correspondence between the feature [labial] and a labial closing gesture, between the feature [nasal] and a velum opening gesture, etc., is evident.



(3)



(4)



(5)

Despite this close correspondence, the representations can not be collapsed. The different way that timing is expressed in the two representations makes features appropriate for expressing lexical contrasts and categorical alternations, and gestures appropriate for expressing gradient processes. Consider two simple examples that illustrate this point: the first example deals with lexical contrasts, the second with phonological rules.

Compare the gestural and autosegmental representations of a labiovelar stop in Figure 7. Figure 7A shows an autosegmental representation, Figure 7B a gestural representation. Both representations involve two articulators: lips and tongue body (labial and dorsal). The most basic difference in the representations is how temporal organization is expressed. Features do not have specific durations, and the only temporal relations that can be defined in this representation are linear precedence and an unspecified amount of overlap (see Bird & Klein, 1990; Sagey, 1988). Association among features is expressed by linkings to abstract hierarchical nodes. Features associated to a given hierarchical node (and not on the same tier) are assumed by the phonology to overlap in time, but the degree of overlap remains unspecified. Thus, there can be no contrast between two kinds of labiovelars that differ only in timing: labial closure followed by dorsal (/b͡g/) and dorsal closure followed by labial (/g͡b/).

In contrast, gestures have inherent extent in time. Precise overlap relations can and must be specified, and are crucial for describing articulatory patterns. Organization among gestures is expressed through the direct specification of timing relations (phasing) between two or more gestures, not through linkings to abstract nodes. In the theory of Articulatory Phonology, phase relations can in themselves serve as the basis of phonological contrast: for example, the difference between aspirated and unaspirated stops is encoded in the phase relations between the glottal and oral gestures (Goldstein & Browman, 1986). Given that several different phasings are possible between the labial and tongue body gestures in a labiovelar stop, there is no reason the phase differences could not serve as the basis of lexical contrast. Thus Articulatory Phonology predicts a possible contrast between /b͡g/ and /g͡b/. The same holds true (as Clements, 1992 points out) of phasings between glottal opening and oral gestures. A gestural representation predicts many possible contrasts, for example between pre-aspirated, unaspirated, and post-aspirated stops, when in fact no language has more than a two-way contrast.

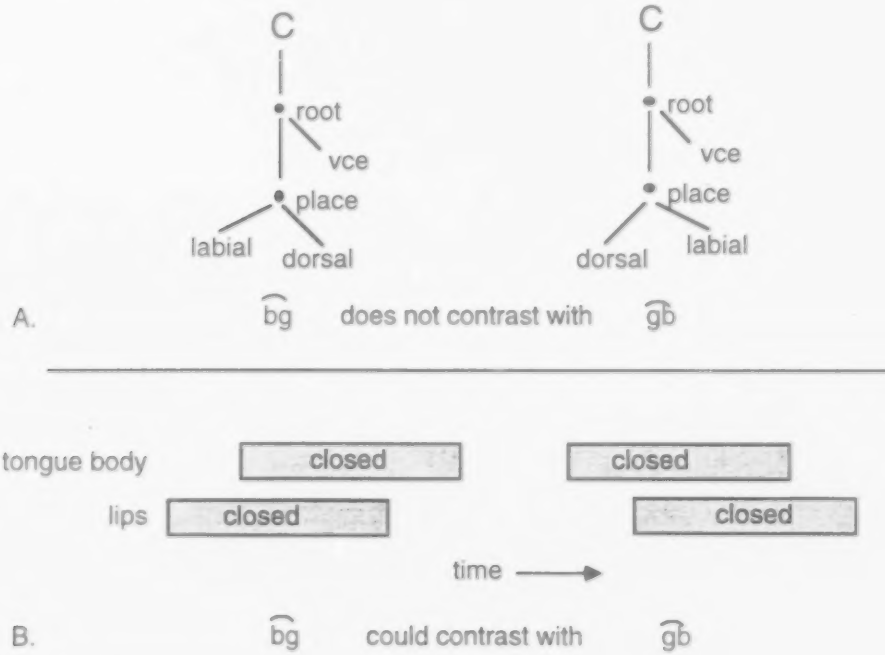


Figure 7. **Representations of a labio-velar stop. A. Featural. B. Gestural.**

In fact, /g͡b/ (dorsal closure first, labial closure second) is almost invariably chosen as the articulatory organization (Connell, 1991; Maddieson & Ladefoged, 1989). As Connell (1991) points out, the phase relations among the different component gestures are crucial for understanding the phonetic behavior of these stops, as well as their diachronic development. Therefore, a phonetic representation must be able to express the asymmetry of the dorsal and labial gestures. But any phonological representation that has the power to express that timing relationship makes wrong predictions about possible synchronic phonological contrasts. It also makes wrong predictions about possible phonological rules.

Consider two rules of nasalization, one categorical, the other gradient. Categorical nasalization can be represented as spreading of the feature [nas] from consonant to vowel (Figure 8A). Because there is no way for a feature to spread only part way from one root node to another, the result is categorical assimilation. To express partial nasalization, explicit timing must be taken into account.

Figure 8B shows a gestural representation of partial nasalization. Because gestures have extent in time, specific points in one gesture are timed with respect to specific points in another gesture. In this gestural score, maximum velic opening is timed to occur at the beginning of the tongue tip gesture for the final /n/. (Krakow, 1989 found this timing relation to hold of nasalized vowels in American English.) In order to achieve this timing with respect to the consonant, the velum opening gesture must begin during the vowel, resulting in partial nasalization. A gestural representation can also capture complete nasalization, by specifying a different timing relation. The vowel and the velum opening could begin at the same time, so that nasalization extends throughout the vowel.

The gestural approach thus uses the relation of overlap in time to describe complete as well as partial processes. Yet specific timing is unnecessary for the description of synchronic phonological rules. In the lexical component, where all rules are categorical, and where reference to abstract hierarchical nodes like the root node is necessary, a theory that allows specific temporal relations to be manipulated is too powerful. While specifying timing relations among gestures is appropriate for gradient rules, categorical rules require the all-or-nothing, plus-or-minus specification that feature spreading provides. However, because the two representations are so similar in respects other than timing, the mapping between them is straightforward. Gestural scores can be seen as feature trees with elaborated timing information, or feature trees can be seen as gestural scores underspecified for temporal relations.

This paper has argued for two different representations for lexical and post-lexical palatalization: an autosegmental featural representation for the categorical lexical rule and a gestural representation for the gradient post-lexical rule. Articulators form the basis for both representations. They differ principally in the kind of temporal information that is available for manipulation: specific extent in time for gestures, only simultaneity and precedence, expressed in terms of linking to hierarchical nodes, for features. This simple correspondence between features and gestures leads to a simple correspondence between categorical and gradient rules.

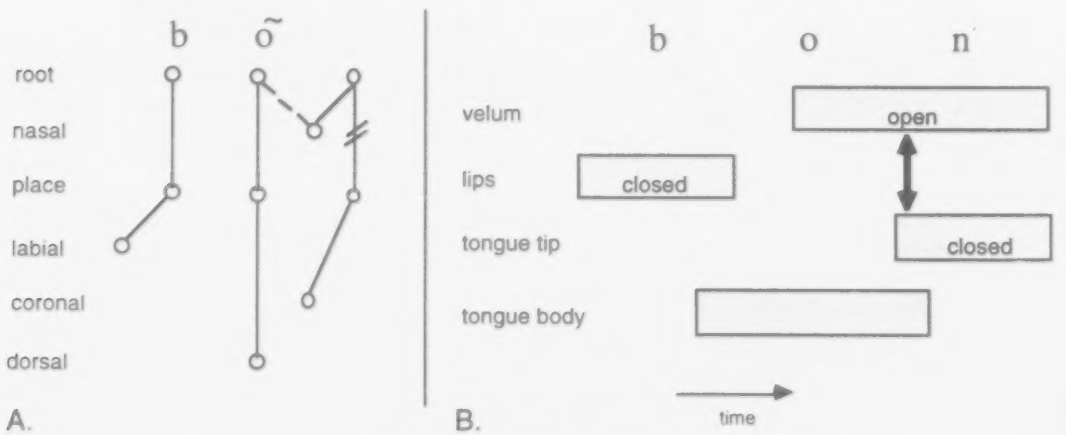


Figure 8. **Two rules of nasalization. A. Categorical. B. Gestural.**

REFERENCES

Bird, S., & Klein, E. (1990). Phonological events. *Journal of Linguistics, 26,* 33-56.

Borowsky, T. J. (1986). *Topics in the lexical phonology of English.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Broselow, E., & Niyondagara, A. (in press). Morphological structure in Kirundi palatalization: Implications for feature geometry. In F. Katamba (Ed.), *Studies in Inter-lacustrine Bantu Phonology.* Cologne: Afrikanistiche Arbeitspapiere.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook* 3: 219-252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6,* 201-251.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 341-376). Cambridge: Cambridge University Press, .

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49,* 155-180.

Catford, J. C. (1977). *Fundamental problems in phonetics.* Bloomington: Indiana University Press.

Clements, G. N. (1976). Palatalization: Linking or assimilation? *CLS : Proceedings of the Chicago Linguistics Society, 12,* 96-109.

Clements, G. N. (1991). Place of articulation in consonants and vowels: A unified theory. In B. Laks & A. Rialland (ds.), *L'Architecture et la Geometrie des Representations Phonologiques.* Paris: Editions du C.N.R.S.

Clements, G. N. (1992). Phonological primes: Features or gestures? *Phonetica, 49,* 18-193.

Connell, B. (1991). Accounting for the reflexes of labial-velar stops. *Proceedings of the XIIth International Congress of Phonetic Sciences* 3: 110-113.

Goldstein, L., & Browman, C. P. (1986). Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics, 14,* 339-342.

Hardcastle, W. J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication, 4,* 247-63.

Hardcastle, W. J., & Clark, J. E. (1981). Articulatory, aerodynamic, and acoustic properties of lingual fricatives in English. *Phonetics Laboratory University of Reading Work in Progress, 3,* 51-79.

Hardcastle, W. J., Gibbons, F., & Nicolaidis, K. (1991). EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics, 19,* 251-256.

Hardcastle, W. J., & Roach, P. J. (1977). An instrumental investigation of coarticulation in stop consonant seqences. *Phonetics Laboratory University of Reading Work in Progress.*

Hume, E. (1990). Front vowels, palatal consonants, and the rule of umlaut in Korean. *NELS: Proceedings of the North East Linguistics Society* 20: 230-243.

Hume, E. (1992). *Front vowels, coronal consonants, and their interaction in non-linear phonology.* Unpublished doctoral dissertation, Cornell University.

Keating, P. (1988). Palatals as complex segments. *UCLA Working Papers in Phonetics, 69,* 77-91.

Krakow, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures.* Unpublished doctoral dissertation, Yale University.

Maddieson, I., & Ladefoged, P. (1989). Multiply-articulated segments and the feature hierarchy. *UCLA Working Papers in Phonetics, 72,* 116-138.

Marchal, A. (1988). Coproduction: Evidence from EPG data. *Speech Communication, 7,* 287-295.

McCarthy, J. J. (1988). Feature geometry and dependency: A review. *Phonetica, 45,* 84-108.

McCarthy, J. J. (in press). The phonetics and phonology of Semitic pharyngeals. In P. Keating (Ed.), *Papers in Laboratory Phonology III.* Cambridge: Cambridge University Press.

Ní Chiosáin, M. (1991). *Topics in the phonology of Irish.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Ohman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39,* 151-168.

Padgett, J. (1991). *Stricture in feature geomtery.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Perkell, J. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study.* Cambridge, MA: MIT Press.

Recasens, D. (1984). Timing constraints and coarticulation: Alveo-palatals and sequences of alveolar + /j/ in Catalan. *Phonetica, 41,* 125-139.

Recasens, D. (1990). The articulatory characteristics of palatal consonants. *Journal of Phonetics, 18,* 267-280.

Sagey, E. C. (1986). *The representation of features and relations in nonlinear phonology.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Sagey, E. C. (1988). On the ill-formedness of crossing association lines. *Linguistic Inquiry, 19,* 109-118.

Shattuck-Huffnagel, S., Zue, V. W., & Bernstein, J. (1978). An acoustic study of palatalization of fricatives in American English. *Journal of the Acoustical Society of America, 64,* S92.

Shibata, S., Ino, A., Yamashita, S., Hiki, S. Hiritani, S., & Sawashimi, M. (1978). A new portable type unit for electropalatography. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 12,* 5-10.

Zsiga, E. C. (1993). *Features, gestures, and the temporal aspects of phonological organization.* Unpublished dissertation, Yale University.

Zsiga, E. C. (1994). Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics, 22,* 121-140.

Zue, V., & Shattuck-Huffnagel, S. (1980). Palatalization of /s/ in American English: When is a /S/ not a /S/? *Journal of the Acoustical Society of America, 67,* S27.

FOOTNOTES

*To appear in B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology IV.* The results and discussion presented in this paper are condensed from Zsiga, 1993.

†Department of Linguistics, Georgetown University.

Haskins Laboratories Status Report on Speech Research
1994, SR-115/116, 81-112

The Discriminability of Nearly Merged Sounds*

Alice Faber and Marianna Di Paolo†

In a near merger, speakers produce two contrasting words differently without being able to reliably discern the contrast in their own speech or in the speech of others. Acoustic measurements typically reveal small differences between the elements of near merged minimal pairs, along several acoustic dimensions. This paper argues that statistical evaluation of the potential distinctiveness of these near merged elements must take simultaneous account of all these dimensions. For that reason, discriminant analysis was used to assess the differences between near merged /il-ɪl/, /el-ɛl/, and /ul-ʊl/ for five Utah speakers. In contrast with independent univariate Analyses of Variance of F1, F2, f_0, and spectral slope, the multivariate discriminant analyses suggest that all three contrasts are preserved by all five speakers. However, homophones like *heel* and *heal* were not distinguished by the discriminant analyses. Discriminant analysis is thus a powerful technique for assessing whether a reliable basis exists for the claim that two potentially contrastive items are in fact distinctive.

1 INTRODUCTION

The phenomenon of *near merger* in language change has gained increasing attention in recent years but it is still not very well-understood. In near mergers, two sounds which were originally distinct appear at one synchronic language stage to have merged completely; however, at a later stage the two sounds are again distinct. The paradigm case of such a near merger is that of Early Modern English reflexes of Middle English */ɛ̄/ as in *meat*, which first appears to have merged with reflexes of */ā/ as in *mate*, but which in Modern English has merged instead with reflexes of */ē/ as in *meet* (Labov, 1974; Harris, 1985; Faber, Di Paolo, & Best, 1994). The most plausible diachronic account, then, is that the original merger was an illusion (similarly, Nunberg, 1980). Indeed, in present-day near mergers, as reviewed by Labov, Karen, & Miller (1991) and, more recently, Labov (1994: Part C), it is often the case that speakers are demonstrably unaware of small but reliable acoustic differences that they equally demonstrably produce. As Labov notes, near mergers present a challenge for linguistic theory, in that it is unclear how such seemingly imperceptible differences could be learned. Yet, if such differences are learned—and if they are observed, it must be that the speakers for whom they are observed did in fact learn them—they must have been perceptible to these speakers at some point in the process of language acquisition (see Faber, Di Paolo, & Best [ms] for details). In the present paper, our interest lies in development of an appropriate framework to describe the small but reliable differences among speech sounds of the sort that characterize near mergers.

The first author acknowledges support from NIH grants DC–00403 to Catherine Best and DC–00016 and HD–01994 to Haskins Laboratories during preparation of this paper. The second author acknowledges support from an Eccles Fellowship in the Humanities (University of Utah) during preparation of the final manuscript. We thank Drs. Marvin Hansen and Lynn Alvord of the Department of Communication Disorders, University of Utah for providing recording facilities and Shari Kendall for assistance with data analysis. Some of the material in this paper was presented in a Symposium of the Speech Communication Group, Research Laboratory of Electronics, MIT and at the 1992 meeting of the Linguistic Society of America, and this paper has benefited from feedback following those presentations. We are grateful to Len Katz for extensive discussion of the statistical techniques used in this paper, and to various colleagues at Haskins Laboratories, at the University of Utah, and elsewhere for discussion of the ideas herein contained. Special thanks are due Cathi Best for her many contributions to the ideas presented here. We hereby absolve all of these people for any misuse we might have made of their ideas and suggestions.

In an earlier study (Di Paolo & Faber, 1990), we examined the near-merger of tense and lax vowels before /l/ in younger speakers in the Salt Lake Valley of Utah.[1] For many, although not for all, speakers in this area, formant differences between the vowels in, for example, *heel* and *hill* are at best minimal, at least in formal word list style (laboratory speech).[2] Neither are there consistent differences in duration or in f_0. There are, however, consistent albeit small differences in spectral slope (amplitude level of f_0 (L0) minus amplitude level of F1 (L1)), which we also refer to as VQI). The tense vowels have more prominent f_0 relative to F1 than do the lax vowels. Just as differences in formant frequency reflect articulatory differences in the configuration of the supralaryngeal vocal tract, differences in spectral slope in vowels with the same formant frequencies reflect differences in laryngeal configuration (Ladefoged, 1983; Stevens, 1988, 1989). The tense vowels are breathy and the lax vowels are creaky. In other words, the glottis is open for a larger proportion of the duty cycle in the tense vowels than in the lax vowels. Our suggestion was that these differences in spectral slope suffice to distinguish tense from lax vowels before /l/ in Utah.

Like most studies of vowel acoustics, our previous study simply equated vowel distinctiveness with distinctiveness along a single measurable dimension. This is not a necessary equation. If we posit a multidimensional vowel space whose axes represent, perhaps, F1, F2, f_0, spectral slope, and duration, it is possible to imagine two vowel nuclei occupying distinct regions of this space but not being well separated along any one of its axes. In such a case, the two vowels would be distinct; however, ordinary analysis techniques, techniques which treat the multiple dimensions of the vowel space in quasi-independent fashion, would not uncover this distinctiveness. In the current study, we use *discriminant analysis*, a statistical technique for assessing the extent to which multiple parameters serve to distinguish groups of items (Klecka, 1980). This technique has previously been applied to linguistic data by Port and Crawford (1989), in a study of German final devoicing; Sussman (1991), Sussman, McCaffrey, and Matthews (1991), and Fowler (1994) in studies of English stop consonant place of articulation; and, Johnson, Ladefoged, and Lindau (1993) in a study of vowel production. Port & Crawford measured vowel duration, final stop closure duration, stop burst duration, and number of glottal pulses in the final stop closure in three minimal pairs differing in underlying stop voicing produced by five German speakers. In separate ANOVAs for each measure, only stop burst duration varied significantly according to underlying stop voicing. Nevertheless, the optimal discriminant function utilized *all* of the measured variables with the exception of number of glottal pulses in the final stop closure. That is, variables that by themselves do not differentiate the two categories—final "voiced" and "voiceless" stops—do contribute to a multidimensional differentiation of the categories. As Port & Crawford stress, this simultaneous reference to multiple weak cues may provide a better analog to actual speech perception than does traditional analysis in terms of single strong cues.

In the present study we extend our previous research on the distinction between tense and lax vowels before /l/ in Utah English. In our previous research (Di Paolo & Faber, 1990), we had suggested, on the basis of acoustic measurements and a perceptual labeling study, that small differences in spectral slope suffice to distinguish cognate tense and lax vowels before /l/. In the present study, we assess multidimensional vowel contrast by means of discriminant analysis. In the earlier study we had included no cases in which contrast was *not* expected; we therefore had no way of determining whether our techniques were too sensitive. In the present study, we therefore include a number of pairs of homophones, in order to provide a baseline of clear lack of contrast.[3] This inclusion of homophones will provide us with a criterion for assessing phonetic distance. If *heel* and *hill* are acoustically more distinct than *heel* and *heal*,[4] this will lend support to a model of sound change in which small but significant differences serve to preserve a contrast while some acoustic parameters are changing. If, on the other hand, the distinction between *heel* and *hill* is comparable to that between *heel* and *heal*, such a view of sound change is not supported.

In addition, our previous study had treated F1, F2, f_0, and spectral slope as independently controlled acoustic parameters, when, of course, they are not. While F1 and F2 may each reflect different aspects of the vocal tract configuration during the articulation of a vowel, a single set of articulatory maneuvers, involving tongue and lip position, controls both, simultaneously. Likewise, the amplitude of F1 (L1) may depend both on its frequency and on the frequency of higher formants (Fant, 1956), so a finding of significant variation in both F1 frequency and in spectral slope (L0-L1) may reflect this dependency rather than variation in the amplitude of f_0 (L0) resulting from variation in laryngeal configuration. If laryngeal configura-

tion is of interest, as it is in the present study, two solutions to this non-independence problem are available. The first involves attempting an independent measure of glottal configuration, either through inverse filtering of the speech signal (Javkin, Antoñanzas-Barroso, & Maddieson, 1987; Löfqvist, 1991) or electroglottography. These techniques are most often used to study speaker-characteristic, pervasive voice qualities rather than time-varying phonological characteristics. In particular, inverse filtering is inappropriate for vowels in which the F1 resonance is low enough in frequency to include the fundamental, and, as a result, inappropriate for study of potential laryngeal configuration differences in high vowels. Likewise, electroglottography (measurement of the changing rate of electrical current transmission across the glottis as the vocal folds open and close) is very sensitive to correct electrode placement, which is easiest to achieve in adult males with very little neck fat (Colton & Conture, 1990); in addition, it is potentially sensitive to variation among vowels in overall larynx height. The second solution to the non-independence problem is a statistical one. Rather than analyzing each parameter independently, discriminant analysis provides a global treatment. The independent contribution of each variable to the overall set of discriminant functions can be assessed. If two variables are perfectly correlated, the second variable makes no independent contribution to the discriminant functions, over and above that made by the first. Thus, to the extent that both F1 and spectral slope contribute to a set of discriminant functions, it can be assumed that some component of the variance in spectral slope is not correlated with F1, and *ex hypothesis* reflects underlying laryngeal configuration.

2 Methods

Eight subjects, five from Utah, and, for purposes of comparison, three from Connecticut, read eight randomizations of the word list in Table 1. Utah subjects were recruited from introductory Linguistics classes at the University of Utah, and were paid $10.00 for their participation. Connecticut subjects were recruited through acquaintance networks at Yale and Wesleyan Universities; Wesleyan subjects were paid $6.00 for their participation, but the Yale subject, a research colleague, was not paid. For the Utah subjects (but not the Connecticut subjects) laryngeal vibration was recorded directly, via a small accelerometer attached with adhesive on the external neck surface opposite the thyroid lamina. These laryngeal signals were not analyzed in the present investigation. Further demographic details about the subjects are given in Table 2. Filler words were used at the beginning and end of each column, and subjects were instructed to read slowly, with a two-beat pause after each word. All material was recorded in sound-treated rooms on a cassette tape recorder. In all, each speaker produced 440 tokens. Tokens for four Utah subjects and for all three Connecticut subjects were digitized at a 10 kHz sampling rate (12 bit quantization, with preemphasis) using the Haskins Laboratories PCM system. Tokens for the fifth Utah subject were digitized at a 8 kHz sampling rate (16 bits quantization, with preemphasis) using an Audiomedia A-to-D board on a Macintosh IIsi microcomputer at the University of Utah.[5] Acoustic measurements were made at three points during each vowel: an early point, the approximate midpoint, and a late point. An effort was made to avoid formant transitions out of and into the adjacent consonants, insofar as possible. With the /l/-final words, this was, of course, impossible, since the influence of the /l/ may extend through the entire vowel; all third measurements in /l/-final words were made in the vocalic portion of the word, and not in the /l/. In addition, one of the Connecticut speakers, C4, had extremely long transitions into /d/; these transitions started in some instances as early as the vowel midpoint, so many of his late measurements reflect the influence of the following /d/. We measured four parameters: F1, F2, f_0, and VQI (spectral slope: L0-L1), for a total of 5280 data points per speaker.[6]

Table 1. *Words used in current study.*

heed	heal	food	fool
he'd	heel	poop	pool
peep	he'll	hood	full
	peal	cook	pull
	peel	hoed	hole
hid	hill	pope	whole
pip	pill		pole
hayed	hale		poll
tape	hail	HUD	hull
	pail	pup	cull
	pale	hawed	hall
head	hell	talk	haul
pep	pell		pall
pap	pal		Paul
had	Hal	cod	Col
		pop	pol
			Sol

Table 2. *Subject characteristics.*

C1:	female, late teens; Middletown (central Connecticut)
C2:	female, late 20's; Stratford (southwest Connecticut)
C4:	male, early 20's; Branford (south central Connecticut)
U5:	male, late 20's; Ephraim (Central Utah)
U6:	female, early 20's; Salt Lake City
U8:	female, mid 30's; Davis County (north of Salt Lake City)
U9:	female, late teens; Salt Lake City
U12:	male, early 20's; Salt Lake City

For each speaker, the following analyses took place. Each parameter served in turn as the dependent variable for an Analysis of Variance, with Vowel Identity and Final Consonant serving as independent, between-tokens factors and Measurement Location as a repeated, within-token factor. (In these analyses, all factors were fixed, except, of course, for token.) Given the highly significant main effects for Final Consonant and Measurement Location, and the particular interest of this study in vowel distinctions before /l/, separate ANOVAs were then performed on the subset of /l/-final words at each measurement location, using Vowel Identity as the independent variable. Then, each parameter was standardized, using *z*-transforms.[7] The transformed parameter values served as input to three separate discriminant analyses for each speaker, the Single Speaker analyses. In the first of these analyses, target Word was the grouping variable, and the 12 acoustic measures (4 parameters × 3 measurement locations) were the dependent variables. In the second analysis, VC Rhyme (e.g., /id/, /ul/, /op/, etc.) was the grouping variable, and the 12 acoustic measures were, as in the first analysis, the dependent variables. The third analysis was like the second, except that discriminant functions were calculated for the /l/-final words only. Finally, additional Word and VC Rhyme analyses were performed in which productions by each of the Utah female subjects U6, U8, and U9 were classified according to discriminant functions derived from productions by the other two Utah females and productions by C2, the Connecticut female from whom we have complete data; these are the Two Speaker analyses. The Two Speaker analyses were restricted to the female subjects because of suggestions (e.g., Henton, 1992; Johnson, 1989) that female speakers' vowel spaces are proportionally larger than males', even when formant frequencies are normalized to the same range of the frequency scale. All discriminant analyses were done using BMDP program 7M (Dixon, 1988), with all dependent variables forced into the discriminant function.[8]

Discriminant analysis constructs an *n*-dimensional coordinate space, corresponding to the input variables (see Klecka [1980] for details).[9] This space is rotated so as to maximize the amount of variation on a minimum number of axes, and form a space of smaller dimension than the original. These derived axes are referred to as canonical variables or discriminant functions. The mean coordinates for each group within the rotated coordinate space are calculated.[10] The overall significance of the discriminant analysis reflects the extent to which the input groups occupy disparate regions of the coordinate space. Thus, the more distinct and non-overlapping the input groups actually are, the higher the significance level. The geometric distance in the coordinate space between each token and every group mean is calculated, and the token is assigned to the group to which it is closest, regardless of its original group membership. In this calculation, the token being classified is excluded from the group means to which it is being compared; thus, our classifications are made by the jackknife method. The extent to which specific tokens are classified into their original groups allows, in the present instance, assessment of homophony, and even potential sound change. Secondly, for each pair of group means, a partial F-value is calculated, on the basis of which the likelihood can be computed that given pairs of groups are distinct in the derived coordinate space. This assessment complements the one arrived at through examination of the classification matrices. In cases like the present, in which the number of pairwise comparisons proliferates (1485 in the case of the Word analyses, 508 in the case of the VC Rhyme analyses, and 55 in the case of the Rhyme, /l/ words only analysis), extreme caution must be exercised in interpreting these partial F values, due to the increased likelihood of Type I error. Finally, it provides partial F-values for each of the input variables on the basis of which the relative strengths of their contributions to the derivation of the canonical variables can be ranked.

3 Results

Prior to presenting the results of the main analyses, we will first describe the speakers' productions in fairly conventional terms, in section 3.1. In section 3.1 we also present the results of our

univariate analyses of variance. After outlining the sorts of results that might be expected in the Single Speaker discriminant analyses, in section 3.2 we present the results of these analyses. The results of the analyses with target word as the grouping factor (section 3.2.1) show that the acoustic measures that we made are in the aggregate sufficient to distinguish among words with different vowel nuclei, even though they do not distinguish between homophones. They also differentiate words with nearly merged nuclei. The analyses with VC Rhyme as the grouping factor (section 3.2.2) confirm that our acoustic measures distinguish among different vowels in general, and between the nearly merged Rhymes /ɪl/-/ɪl/, /ʊl/-/ʊl/, and /el/-/ɛl/ in particular. In addition, examination of the canonical variables derived by these discriminant analyses provides some indication of how these contrasts are implemented. Then, in section 3.3, we discuss the Two Speaker analyses, which show that, in general, the Utah speakers' productions are better classified by discriminant analyses trained on the other Utah speakers' productions than by analyses trained on the Connecticut speaker's productions.

3.1 Univariate Analyses

Figure 1 shows the vowels before /d/ for C1 in F1/F2 space, and Figure 2 shows the same vowels for U6. Several features of these vowel spaces are worth noting. For both of these young female speakers, /u/ and /ʊ/ have much higher F2 than would be expected on the basis of Peterson and Barney (1952) and other similar studies.[11] For both speakers, F2 decreases throughout much of /u/ and increases throughout /ʊ/. And for both speakers, /o/ decreases throughout in both F1 and F2, although there is more formant movement for the Connecticut speaker than for the Utah speaker. The single most salient difference between the two speakers lies in the low vowels. For the Connecticut speaker in Figure 1, /ɑ/, /ɔ/, and /ʌ/ are quite distinct, whereas for the Utah speaker in Figure 2, /ɑ/ and /ɔ/ have similar but not overlapping formant tracks[12] and /ʌ/ starts in the same region, rising to approach /ɛ/.

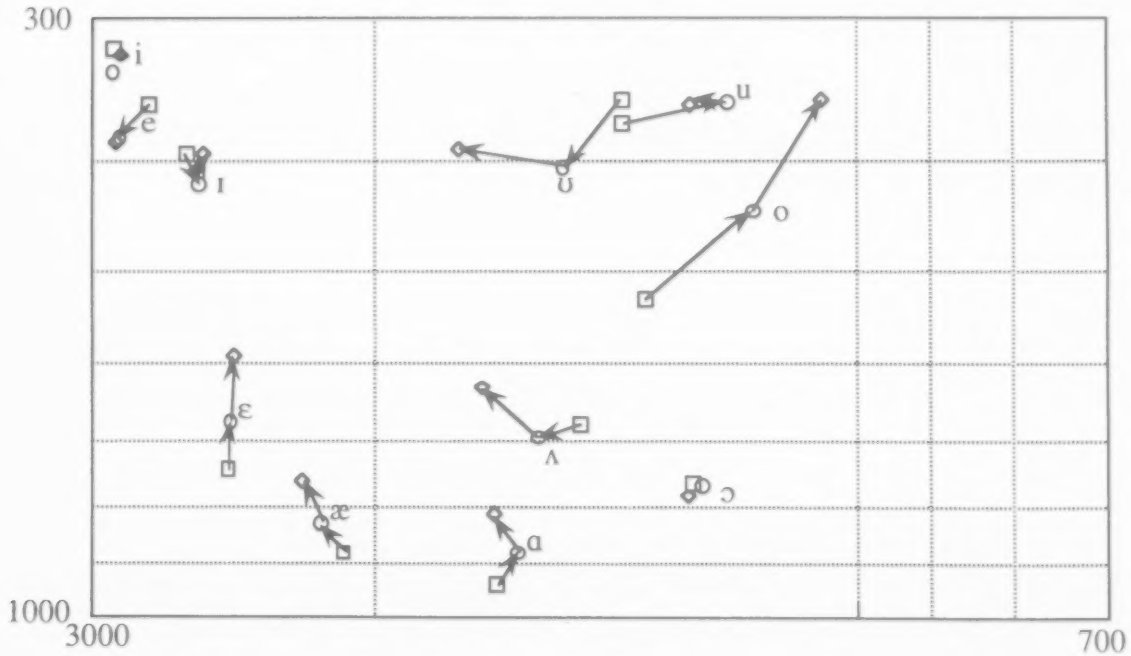


Figure 1. **Formant tracks for C1's vowels before /d/.**

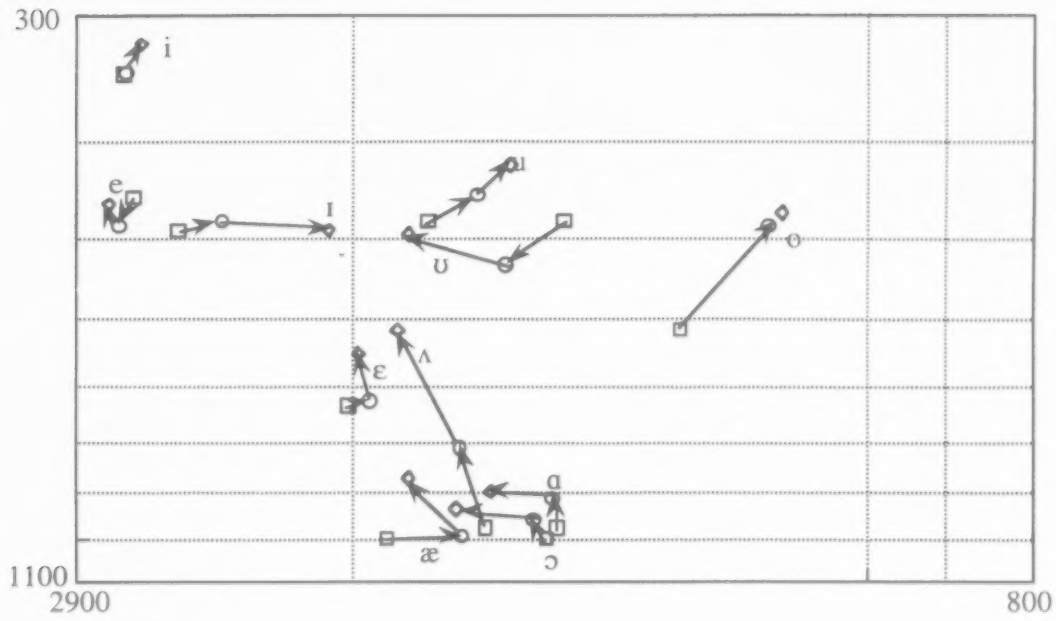


Figure 2. **Formant tracks for U6's vowels before /d/.**

Figures 3 and 4 show the same speakers' vowels before /l/. Here again, there are several features common to the two dialect areas, namely the tendency to lower F2 for vowels before /l/ and the fact that /u/ and /ʊ/ have much lower F2 before /l/ than in other contexts. The overlap of /ʌl/ and /ʊl/ in Figure 3 is less typical, and may be a conservative feature.[13] The three-way approximation of /ol/, /ʌl/ and /ʊl/ in Figure 4 is typical of some younger speakers in the Salt Lake Valley (see below, section 4.2, for further discussion). Similarly, the fact that /ɪl/ starts with *higher* F2 than does /il/ is not uncommon in younger Utah speakers. Like many of the speakers described in Di Paolo & Faber (1990), this speaker has fairly parallel formant tracks for /il/ and /ɪl/. She differs from other younger speakers in maintaining a clear distinction between /ul/ and /ʊl/, perhaps because /ul/ has an unusually high F2 for this context. She is unusual also in having /el/ in the same area as /il/ and /ɪl/: some speakers have /el/ in the same area as /ɛl/; others have /el/ lower than /il/ and /ɪl/, but still distinct from /ɛl/, which in turn heavily overlaps with /æl/.

Overall, our ANOVAs for all subjects show significant differences among vowels in all of the parameters measured. In addition to the formant differences already discussed, all speakers clearly have vowel-related differences in f_0, regardless of following consonant. The speakers differ, however, in the range of f_0 variation and in the relative ranking of vowels in f_0. There are also significant differences among vowels in VQI, but it is unclear how to interpret these differences, since, as already noted, the amplitude of F1 (L1) in a particular instance depends partly on F1 frequency, and so VQI (L0-L1) will depend on F1 frequency as well as on underlying laryngeal configuration. However, when the speakers' vowels before /d/ are ranked from breathy to creaky, the order does not match the order predictable on the basis of Fant's (1956) L1 levels. Thus, our variation in VQI results from variation in L0 as well as from variation in L1.

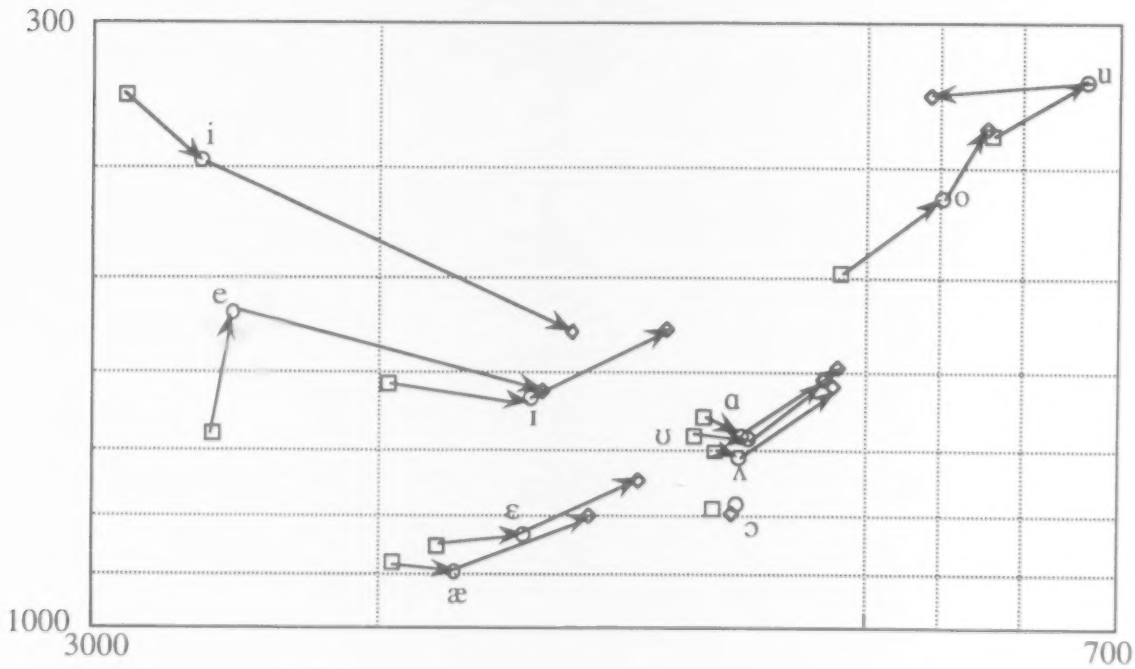


Figure 3. **Formant tracks for C1's vowels before /l/.**

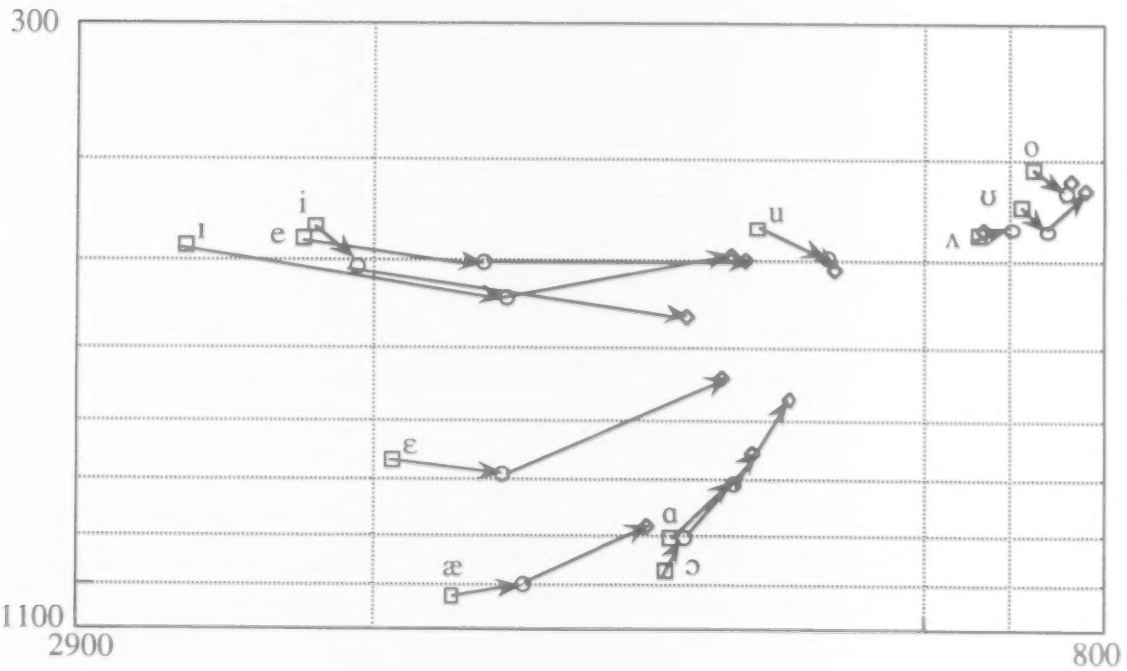


Figure 4. **Formant tracks for U6's vowels before /l/.**

Table 3. *Summary of Bonferroni/Dunn test for selected contrasts from ANOVA's for /l/-final words at three measurement points. Squares left blank indicate p > .05. The first row in each cell reports on the early measurement point, the second row on the middle measurement point, and the third row on the late measurement point. Because of recording problems, only formant analysis is available for C1.*

		U5	U6	U8	U9	U12	C2	C4	C1
/i/-/ɪ/	F1	<.0001 <.0001 <.0001			<.0001	<.0001 .0002	<.0001 <.0001	<.0001 <.0001	<.0001 <.0001
	F2	<.0001 <.0001 <.0001	<.0001		<.0001 <.0001	<.0001 <.0001	<.0001 <.0001	<.0001 <.0001	<.0001 <.0001 .0003
	f0						<.0001		— — —
	VQI	<.0001 <.0001		.0005			<.0001 <.0001		— — —
/e/-/ɛ/	F1	<.0001 <.0001 <.0001	<.0001 <.0001		<.0001 <.0001	<.0001	<.0001 <.0001	<.0001 <.0001	<.0001 <.0001 .0004
	F2	<.0001 <.0001 <.0001	<.0001	.0002	<.0001 <.0001 <.0001	<.0001 <.0001	<.0001 <.0001	<.0001 <.0001 .0002	<.0001 <.0001
	f0						<.0001		— — —
	VQI						<.0001		— — —
/u/-/ʊ/	F1	<.0001 <.0001			<.0001 <.0001	<.0001	<.0001 <.0001	<.0001 <.0001	<.0001 <.0001 <.0001
	F2		<.0001 <.0001 <.0001		<.0001		<.0001 <.0001	<.0001 <.0001	<.0001 <.0001
	f0	<.0001					<.0001 <.0001		— — —
	VQI	<.0001 <.0001					<.0001 <.0001		— — —

The separate ANOVAs on the vowels in /l/-final words at each Measurement Point, with target Vowel identity as the independent variable likewise show highly significant ($p \leq .0001$) differences among vowels on all four dimensions, but post hoc comparisons (Bonferroni/Dunn) of the specific contrasts of interest, summarized in Table 3, reveal a rather murky picture. For the Connecticut speakers, all three contrasts are maintained. Likewise in Utah, for U5, all 3 contrasts are maintained, and for U8, the oldest speaker in our sample, /il/-/ɪl/ and /el/-/ɛl/ are only distinct in one parameter at one measurement point; /ul/-/ʊl/ are not at all distinct. For U6, U9, and U12, all three are maintained, but in some cases in only one parameter at one measurement point. So, U6's /il/-/ɪl/ are distinct only in F2 at the early measurement point, and her /ul/-/ʊl/ are only distinct in F2 at all measurement points. U12's /ul/-/ʊl/ are distinct only in F1, and only at the first measurement point. In virtually all of these cases, the contrasting vowel pairs differ only in F1 and/or F2. VQI and f_0 systematically differentiate tense-lax cognates before /l/ only for C2, U8, and U5, and the relevant pairs (with the exception of U8's /il/-/ɪl/) also differ in F1 and/or F2.

3.2 Multivariate analyses: Discriminant analysis

Before we describe the results of the discriminant analyses, we must lay out the sorts of results that we should expect. First of all, unambiguous misreadings should be classified in accord with listeners' perceptions. For example, one subject fairly consistently read *peep* as *pep*. These tokens should be classified by the discriminant analysis as *pep* or *head* rather than as *peep* or *heed*. Secondly, we would expect homophones like *peal* and *peel* to be confused with each other as often as they are classified correctly. A priori predictions regarding non-homophones with the 'same' vowel are less clear; while our measurements aimed at avoiding formant transitions out of and into adjacent consonants, we may not always have been successful. Furthermore, it is not clear that the influence of context consonants on vowels is restricted to the transitions (e.g., Sussman, 1991). As a result, it is not clear to what extent *heel* and *peel* will be confused. However, they should be confused with each other more than with other words. In addition, words with different vowels should have different patterns of confusion. Thus, if two words have different patterns of confusion, we can infer that they have different vowel targets. And, given that there is a considerable amount of variation among speakers in the position of vowels before /l/ in formant space, we should not be surprised to find inter-speaker variation in patterns of distinctiveness.

3.2.1 Analyses with word as the grouping factor

All of the above predictions are clearly borne out by the Single Speaker analyses using Word as the grouping factor. All misreadings were classified in accord with listeners' perception, and were excluded from subsequent analyses. Furthermore, on the basis of the unexpected classification of one token of *Hal* as *hill*, we were able to identify and correct a measurement error. Table 4 shows the discriminant analyses' classifications of /l/-final homophones.[14] Overall, tokens of *hail* and *hale*, for example, were classified as *hail* or *hale* ("self" or "homophone") more often than they were classified as *pale* or *pail* ("same rhyme"); they were, in addition, classified as *pail* or *pale* more often than they were classified as *hayed*, *head*, *hill*, *pip*, or *food* ("different rhyme"). The "different rhyme" category includes classification of *heel* as *heed* or as *hill*. In spite of the fact that discriminant analysis shows cognate tense-lax vowel pairs before /l/ to be distinct in both dialect areas, there are different classification patterns for speakers in the area undergoing sound change than in the area not undergoing change. In particular, the larger number of "different rhyme" classifications for Utah speakers than for Connecticut speakers reflects the fact that some Utah tokens of *heel* were classified as *hill*, and *vice versa*.

Table 4. *Summary of Single Speaker target Word discriminant analyses' classifications of members of /l/-final homophone sets, expressed in percent of words in homophone sets receiving a particular classification. Clear speaker misreadings have been excluded.*

Subj	N	Same Rhyme				Dif. Rhyme
		Self	Homophone	Same Rhyme	Total Same	
U5	134	21.64	19.40	44.78	85.82	14.18
U6	136	25.74	20.59	32.35	78.68	21.32
U8	135	25.19	28.89	28.15	82.23	17.79
U9	129	34.88	23.26	23.26	81.40	18.60
U12	133	24.06	26.31	34.59	84.96	15.04
C2	136	28.68	38.24	29.41	96.33	3.68
C4	120	37.50	22.50	32.50	92.50	6.67

Examination of the patterns of classification for individual words reveals that there are, indeed different patterns of confusion for cognate tense and lax vowels. For example, as shown in Table 5, three tokens each of U8's *fool* and *pool* were identified as *full* or *pull*, compared with five tokens each of *full* and *pull*: despite the substantial confusion, the patterns of confusion remain distinct. (The complete classification summaries are presented in Appendices A–C.) For all three vowel locations, the Connecticut speakers' words are somewhat more accurately classified than the Utah speakers' words, and the sorts of "misclassifications" are different.

Table 5. *Classification of words with high back vowels before /l/ from Single Speaker target Word discriminant analyses, for subject U8. Figures represent the number of tokens (out of 8) for which each classification occurred. Rows represent U8's targets, and columns the classifications by the discriminant analysis.*

Classification/ Target	fool	pool	full	pull	other
fool	3	1	1	2	1
pool	2	3	2	1	
full	1	1	2	3	1
pull	1	2	2	3	

For example, both the Connecticut speakers have tokens that are classified at a different general location (for example, high front classified as mid front); however, for Connecticut speakers, tense vowels are always classified as tense, and lax vowels as lax, while for the Utah speakers, there are some instances of tense vowels classified as lax and *vice versa*. For the Utahns, there appears to be no difference among *heel*, *heal*, and *he'll*; all are equally likely to be confused with *hill*, *pill*, or *pip*, as shown in Table 6. However, for one of the Connecticut speakers, C4, *he'll* alone is confused with *hill* and *pill*, perhaps under the influence of *will* in the uncontracted form *he will*.[15]

3.2.2 Analyses with VC rhyme as the grouping factor

Only in the case of U8's /il/ and /ɪl/ do the confusion patterns from the Word analyses illustrated in Table 6 suggest a possible lack of contrast: as many cases of *heel*, *heal*, *he'll*, *peel*, and *peal* as cases of *pill* and *hill* are classified as having /ɪ/. However, the patterns of homophone classification suggest that the question of vowel contrast can better be addressed through the Rhyme discriminant analyses. The Single Speaker analysis by VC Rhymes clearly shows different patterns of confusion for U8's /il/-/ɪl/ (Table 7), as for all other contrasts before /l/ for all four subjects (The complete analyses are summarized in Appendices D and E.). The partial F matrices from the VC Rhyme (/l/-final words) discriminant analyses, paint a similar picture. For both Connecticut subjects and four of the five Utah subjects, the partial F scores show significant contrasts between all three cognate tense-lax pairs ($p \leq .0009$).[16] For the fifth subject, U8, the contrasts between /il/-/ɪl/ and /el/-/ɛl/ are also significant ($p \leq .0009$), while the contrast between /ul/-/ʊl/ is not ($F[12,220]=8.13$; $p = .0301 > .0009$). Nonetheless, the discriminant analyses are compatible with the ANOVAs summarized in Table 3 (Page 88, above); the two analyses agree in suggesting that U8 may preserve fewer contrasts than do the other subjects. We can then conclude that, for our five Utah subjects, contrast is generally maintained between the pairs of cognate tense and lax vowels before /l/. Near homophones, like *heel* and *hill*, have qualitatively different patterns of contrast then do true homophones like *heel* and *heal*. Even though the pairs of near-homophones are acoustically similar enough that they often cannot be distinguished by English speakers from other dialect areas, and even though they may not be distinct on any single dimension, they nonetheless occupy acoustically distinct and perceptually distinguishable regions of a multi-dimensional vowel space. These differences preserve the phonological contrast between the cognate tense-lax pairs before /l/.

Table 6. *Classification of words with high front vowels before /l/ from Single Speaker target Word discriminant analyses, for selected subjects. Rows represent the speaker's targets, and columns the classifications applied by the discriminant analyses.*

Speaker	Classification/ Target	heal	heel	he'll	peal	peel	hill	pill	other
C4	heal	**3**	2		2	1			
	heel	2	**5**			1			
	he'll		1	**1**			3	3	
	peal	1			**6**	1			
	peel	4	3			**1**			
	hill			1			**5**	2	
	pill						2	**6**	
U8	heal		2	3	1	2			
	heel		**4**	2	1		1		
	he'll		3	**3**		1	1		
	peal	1	1		**1**	1	2	2	
	peel				2	**5**		1	
	hill	1	2		4	1			
	pill					2	2	**4**	
U9	heal	**2**		2	1	2			1
	heel	1		3	1	2		1	
	he'll	1	3	**3**	1				
	peal			3	**2**	3			
	peel	1	1		1	**5**			
	hill						**5**	2	1
	pill						3	**5**	

Table 7. *Classification of words with front vowels before /l/ from Single Speaker VC Rhyme discriminant analyses for subject U8. Rows represent the speaker's target, columns represent the classifications applied by the discriminant analysis, and numbers represent the percentage of tokens of a given category that were classified in the way indicated.*

Classification\ Target	N	/ɪl/	/ɪl/	/el/	/ɛl/	/æl/	/æp/	/ɔl/
/ɪl/	40	.60	.40					
/ɪl/	16	.38	.62					
/el/	31			.65	.35			
/ɛl/	16		.06	.25	.69			
/æl/	16					.81	.06	.13

Besides indicating which potential vowel contrasts our speakers implement, discriminant analysis can also provide indirect hints about which of the acoustic parameters that we measured are used to implement these contrasts. The partial F scores just summarized suggest only that the tense-lax pairs for the most part occupy distinct regions of a multidimensional space defined by the measured variables. However, it is relatively straightforward to assess the contributions of the different measured variables to the overall definition of this multidimensional space. These contributions for each of the speakers are ranked in the rows labeled "vars" in Table 8.[17] For each speaker, only the six input variables with the highest F-to-remove values in the Single Speaker VC Rhyme /l/-final words analyses (that is, those variables that in the aggregate make the strongest contributions to the discriminant functions) are listed.[18] The numbers that appear in Table 8 are the coefficients (standardized by group means) by which the first two canonical variables are derived from the input variables. Standardized coefficients are based on standard deviations and tend to range from -1.5 to +1.5. An input variable with a coefficient that has a relatively high absolute value makes a greater contribution to derivation of a canonical variable than does an input variable whose coefficient is closer to zero (Klecka, 1980).

Table 8. *Variables entering into Single Speaker VC Rhyme discriminant functions (/l/-final words only), ranked by F-to-remove score at final step, together with coefficients (standardized by group means) for each input variable in first two canonical variables produced by discriminant analyses. Numbers in parentheses represent, respectively, the proportion of the total dispersion of the data accounted for by the first canonical variable (d1) and the first and second canonical variables (d1 & d2) together. Letters after variables code measurement locations: early, middle, and late.*

C2	var	F1-m	F2-m	F2-e	F1-e	F1-l	VQI-m
(.80593)	d1	.13	-.79	-.67	.31	.08	.25
(.94963)	d2	.69	.16	-.01	.41	.33	-.17
C4	var	F2-e	F2-m	F1-e	F1-m	VQI-m	F1-l
(.74621)	d1	-1.21	.20	.39	-.12	-.28	.20
(.94225)	d2	-.38	.52	.73	.45	.34	.16
U5	var	F1-e	F2-l	F2-e	F1-m	VQI-e	F1-l
(.82634)	d1	.18	-.55	-.73	-.05	-.10	.05
(.97757)	d2	-.69	-.19	-.06	-.35	-.22	-.29
U6	var	F2-e	F2-m	VQI-e	F1-e	VQI-m	F1-m
(.75376)	d1	-.71	-.66	.29	.86	-.32	-.01
(.90924)	d2	.11	-.23	-.02	-.59	-.21	-.42
U8	var	F2-e	F1-e	VQI-e	F2-l	F1-l	F1-m
(.86243)	d1	-1.20	.89	-.57	-.17	.26	-.26
(.98598)	d2	-.03	-.68	.24	.04	-.07	-.36
U9	var	F2-e	F1-e	F1-m	F2-m	VQI-e	VQI-m
(.81907)	d1	-1.04	.54	-.01	-.15	.29	-.04
(.97470)	d2	-.01	-.53	-.64	-.26	-.33	-.23
U12	var	F2-e	F1-e	F2-m	F1-m	f0-e	VQI-e
(.74501)	d1	-.94	-.28	-.23	-.11	.30	-.15
(.95537)	d2	.15	-.73	-.26	-.41	.24	.08

BEST COPY AVAILABLE

It is clear from the coefficients that U6, U9, and U12 only differentiate among vowels before /l/ at the first and second measurement locations. In contrast, U5 and U8, both of whom are relatively conservative with regard to other features (e.g., they both have relatively low F2 for /ud/), have differences throughout the vowels. (These two speakers are also, unlike the other Utah speakers, from outside the Salt Lake Valley.) The two Connecticut speakers, C2 and C4, likewise have differences throughout the vowels. For the Connecticut speakers, only the frequency of F1 late in the vowel (F1-l) contributes to the discriminant functions, whereas for U5 and U8 both F1 and F2 late in the vowel (F1-l and F2-l) make a contribution.

However, the values of the coefficients for these variables in derivation of the first two canonical variables (the first two axes of the coordinate space derived by the discriminant analysis) in Table 8 are not straightforwardly related to the rankings, which are based on the overall analysis. As is evident from the coefficients in Table 8, interpretation of the canonical variables is extremely difficult. There is no straightforward relationship between individual canonical variables and measured acoustic parameters. Nor is a straightforward relationship imaginable between the canonical variables and distinctive feature specifications, either articulatory or acoustic. This difficulty arises in part because measurements of the same parameter at different points in time are treated as independent parameters. Furthermore, the articulatory gestures that produce each particular VC nucleus may each have multiple acoustic consequences.

Table 9 recapitulates the panel from Table 8 concerning subject U9, with the actual coefficients replaced by an indication of their sign and of the rank of the absolute magnitude of each coefficient. Because of the nature of the standardization process, high values for each raw acoustic parameter are positive, and low values are negative; in the case of VQI, values representing relatively breathy phonation are positive and values representing relatively creaky phonation are negative. For subject U9, the input variables that will maximize d1, the first canonical variable, are negative (that is to say, low) F2 at the early and middle measurement points, positive (that is to say, high) F1 early in the vowel, and positive (that is to say, breathy) VQI. D1 will be maximally negative for vowels with high F2, low F1, and creaky VQI.

Table 9. *Variables entering into Single Speaker VC Rnyme discriminant functions for /l/-final words for speaker U9, ranked by F-to-remove score at final step. The sign of the standardized coefficient by which each of the input variables is multiplied in derivation of the first two canonical variables (d1 and d2) is given in the columns headed "sign," and the relative rankings of the magnitudes of the absolute values of the coefficients is given in the columns headed "rank."*

U9	d1		d2	
	sign	rank	sign	rank
F2-early	neg	1		
F1-early	pos	2	neg	2
F1-middle			neg	1
F2-middle	neg	4	neg	4 (tie)
VQI-early	pos	3	neg	3
VQI-middle			neg	4 (tie)

Thus canonical variable d1 defines a vector ranging from high, front, creaky vowels at the negative end to low, back, breathy vowels at the positive end (as shown for U9 in Table 10). Similarly, d2 is maximized by low F1 at the early and middle measurement points, by negative (creaky) VQI at the early and middle measurement points, and by low F2 at the middle measurement point. And it is maximally negative for vowels with high F1, breathy VQI, and high F2. So d2 defines a vector ranging from low, front, breathy vowels at the negative end to high, back, creaky vowels at the positive end.[19] Thus, for U9, d1 involves moving the tongue up and front or down and back from its neutral position, and d2 involves moving the tongue down and front or up and back from its neutral position. These interpretations can be verified with reference to the display of U9's vowels in the bottom left panel of Figure 5. By similar reasoning, the canonical variables for the other subjects can be interpreted. These interpretations are summarized in Table 10.

The differences among subjects reduce to two: 1) Speakers differ as to whether breathiness is associated with the high, front or the low, back range of d1; and, 2) U5's and U8's d2 are qualitatively different from those of the other speakers in that they do not involve the front-back dimension at all. The first difference simply means that for some speakers VQI varies in such a way that it necessarily reflects phonation differences rather than artifacts of F1 frequency variation; for other speakers, however, VQI, as measured here, may or may not reflect phonation differences.[20]

Table 10. *Summary of vectors defined by first two canonical variables (d1 and d2) from Single Speaker VC Rhyme discriminant analyses (/V-final words only) for all speakers.*

	d1		d2	
	negative	**positive**	**negative**	**positive**
C2	hi fro creaky	lo back breathy	lo fro creaky	hi back breathy
C4	hi fro breathy	lo back creaky	lo fro breathy	hi back creaky
U5	hi fro	lo back	lo creaky	hi breathy
U6	hi fro crea > brea	lo back brea > crea	lo back > fro brea	hi fro > back crea
U8	hi fro breathy	lo back creaky	lo creaky	hi breathy
U9	hi fro creaky	lo back breathy	lo fro breathy	hi back creaky
U12	hi fro brea low f_0	lo ba crea high f_0	lo fro > ba low f_0	hi ba > fro high f_0

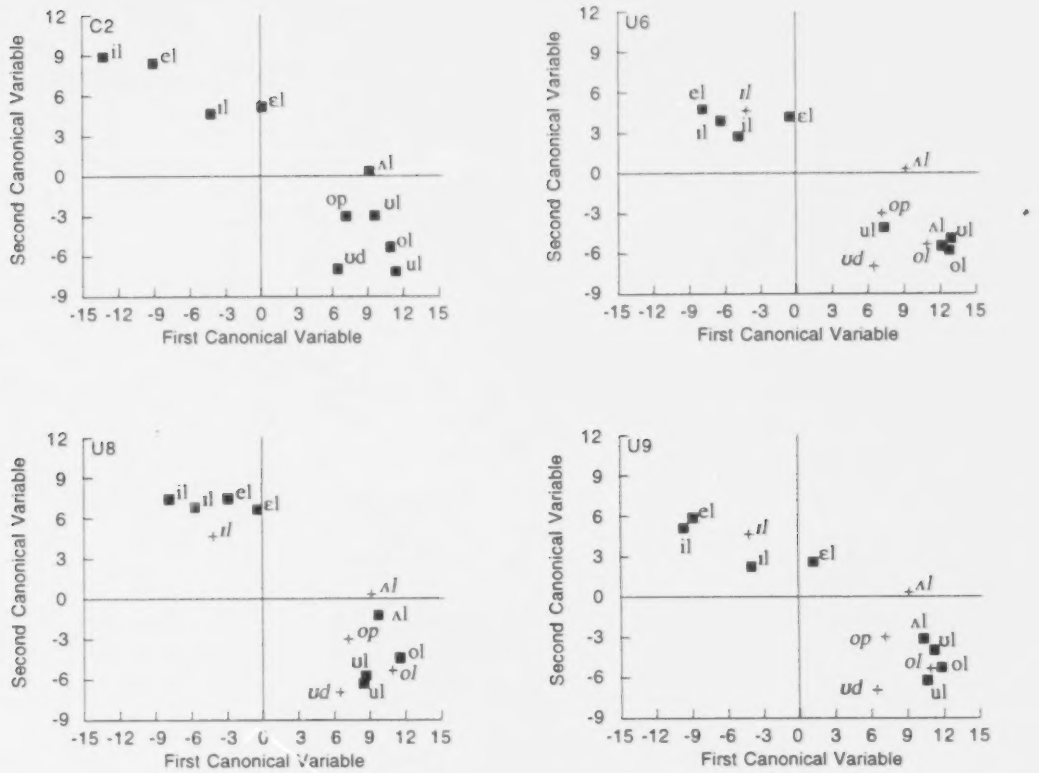


Figure 5. **First two canonical variables from Single Speaker VC Rhyme discriminant analyses for /l/-final words for all Utah subjects. Each plot utilizes a different coordinate system.**

It is worth stressing also that speakers who produce two vowels contrastively may differ as to which parts of the vowels contrast. Our more conservative speakers, whether from Connecticut or Utah, distinguish tense-lax cognate vowels before /l/ at all three measurement locations, while our more innovative speakers distinguish them only at the early and middle measurement locations. Furthermore, our results suggest that the dichotomy between conservative and innovative speakers may need some refinement. According to the ANOVAs summarized in Table 3, U8 makes virtually no distinction between cognate tense and lax vowels before /l/ and, hence, would be classified as innovative with regard to the linguistic phenomenon being studied here. However, as just noted, the discriminant analyses suggest that she does distinguish these pairs, and furthermore, that she distinguishes them in a conservative fashion, in that the distinctiveness is spread throughout the vowel instead of being concentrated in one portion of it.

Returning to the cognate tense-lax pairs, Figure 5 shows the locations of all vowel nuclei before /l/ for each Utah speaker in the Cartesian space defined by the first two canonical variables, d1 and d2 in the VC Rhyme /l/-final analyses, and Table 11 summarizes which canonical variables differentiate which tense-lax vowel pairs. Despite the similarities outlined above, the canonical variables for each speaker are different, and so five different coordinate systems are represented, since each speaker's canonical variables are based on a separate analysis. Each speaker's d1 represents the vector through the point cloud representing that speaker's tokens along which the maximum spread is observed. It is evident from Figure 5 and Table 11 that the subjects differ as to which canonical variables differentiate which tense-lax pair. For C2 and C4, all three pairs are distinguished along both dimensions, d1 and d2, but for none of the Utah speakers is this the case. U5's /il/-/ɪl/ and /el/-/ɛl/ are distinguished by d1 and d2 together, while /ul/-/ʊl/ are distinct only in d2. U6's /el/-/ɛl/ are distinct on both dimensions, while her /il/-/ɪl/ and /ul/-/ʊl/ are distinct only in d1. For U9 and U12, /il/-/ɪl/ are distinguished by both d1 and d2, while /el/-/ɛl/ are distinguished only by d1, and /ul/-/ʊl/ only by d2. Finally, for U8 all three of the relevant pairs are distinct only in d1, and minimally so at that. In summary, then, F1, F2, and VQI contribute to maintenance of the contrasts between cognate tense-lax vowel pairs before /l/ for both the Utah and the Connecticut speakers, although there are qualitative differences between the Utah and the Connecticut implementations of the contrasts.

Table 11. *Implementation of tense-lax vowel contrasts before /l/ by first and second canonical variables from Single Speaker VC Rhyme discriminant analyses of /l/-final words. d1: first canonical variable; d2: second canonical variable.*

Subject	/i-ɪ/	/e-ɛ/	/u-ʊ/
C2	d1>d2	d1>d2	d2>d1
C4	d1>d2	d1>d2	d1>d2
U5	d1>d2	d1>d2	d2
U6	d1	d1>d2	d1
U8	d1	d1	d1
U9	d1>d2	d1	d2
U12	d1>d2	d1	d2

3.2.3 Cross-Speaker Analyses

As already noted, inspection of Tables 8 and 9 above reveals differences between the Connecticut and Utah speakers in the results of the discriminant analyses. Restricting ourselves to the first two canonical variables, d1 and d2, for both Connecticut speakers these variables are derived from F1 throughout the vowel, F2 early in the vowel and mid-vowel, and VQI in mid-vowel. (The two speakers differ, of course, in exactly how the canonical variables are derived from these input variables.) For both Connecticut speakers, the three cognate tense-lax vowel pairs are each distinguished by both of the canonical variables.

The Utah speakers differ from the Connecticut speakers in that for each Utah speaker at least one cognate tense-lax pair is only distinct in one of the canonical variables. Furthermore, different acoustic variables enter into the derivation of the first two canonical variables for the Utah speakers than for the Connecticut speakers. As already noted, these canonical variables are derived only from variables measured early in the vowel or at mid-vowel for U6, U9, and U12. In addition, VQI at mid-vowel is supplemented in derivation of the canonical variables by early VQI for U6 and U9 and replaced altogether by early VQI for U5, U8, and U12.

These differences suggest that discriminant functions based on the Connecticut corpora should be less accurate at classifying the Utahns' words than were the discriminant functions based on the Utahns' own productions. For these reasons, the Two Speaker discriminant analyses were performed for the female Utah speakers, U6, U8, and U9. For each of these analyses, discriminant functions were derived based on Connecticut

female C2's productions, and then the Utah speakers' productions were classified based on these discriminant functions.[21] Because any Two Speaker analysis can be expected to be less accurate than its corresponding Single Speaker analysis, regardless of the dialect(s) of the speakers, additional Two Speaker analyses were performed in which each of the female Utah speakers' productions were classified according to the vowel systems of the other two Utah females. These Utah-Utah analyses serve as a control for the Connecticut-Utah analyses.

In the Single-Speaker analyses for these speakers (Section 3.2.1), the individual words with /il ɪl el ɛl ul ʊl/ had, for the most part, been classified differentially. However, as shown in Table 12, the situation is different for the Connecticut Two Speaker Word analyses.

Table 12. *Pooled classification of Utah females' (U6, U8, and U9) selected /l/-final words from Two-Speaker target Word discriminant analyses. The top part of each panel represents the average of classifications based on the other two Utah speakers and the bottom part classifications based on the Connecticut speaker.*

A. /il/-/ɪl/		heal	heel	he'll	peal	peel	hill	pill	other
< Utah	heal	.15	.06	.21	.08	.13	.19	.06	.12
Subjs.	heel	.13	.21	.31	.04	.06	.06	.10	.09
	he'll	.13	.19	.29	.04	.10	.04	.08	.13
	peal	.08	.08	.06	.02	.16	.21	.21	.18
	peel	.06	.06	.04	.10	.16	.27	.16	.15
	hill	.04	.02	.13	.19	.23	.19	.04	.16
	pill	.02	.04	.06	.06	.27	.31	.02	.22
		heel	hill	pill	hail				
< C2	heal	.21	.71	.04	.04				
	heel	.42	.50	.04	.04				
	he'll	.35	.43	.13	.09				
	peal	.08	.88		.04				
	peel		1.00						
	hill	.08	.88	.04					
	pill		1.00						

B. /el/-/ɛl/		hail	hale	pail	pale	hell	pell	peel	other
< Utah	hail	.35	.08	.17	.17	.02	.02	.04	.15
Subjs.	hale	.20	.09	.09	.35	.04	.07	.02	.14
	pail	.19	.04	.17	.31		.02	.13	.14
	pale	.20	.02	.30	.39	.02		.02	.05
	hell	.04		.10	.02	.25	.46	.06	.07
	pell			.06	.06	.21	.54	.02	.11
		hail	pale	hell	pell	heel	hill	pill	other
< C2	hail	.13	.04	.04		.21	.54	.04	
	hale	.13	.09			.04	.61	.13	
	pail		.13			.04	.71	.08	.04
	pale	.09	.05	.09		.09	.64		.04
	hell			.38	.08		.42		.12
	pell			.50	.17		.21	.04	.08

C. /ul/-/ʊl/		pool	full	hull	hoed	hole	whole	pole	cook	other
< Utah	pool	.10	.10	.15	.04	.06	.04	.08	.33	.10
Subjs.	fool	.10	.13	.08	.19	.17	.04	.15		.14
	pull	.08	.06	.10	.06	.15	.15	.25		.15
	full	.13	.06	.13		.17	.13	.21		.17
		pull	full	hood	whole	other				
< C2	pool	.04	.21	.50	.08	.17				
	fool		.13	.58	.08	.21				
	pull	.13	.17	.17	.29	.24				
	full	.08	.38	.04	.33	.17				

The Utah speakers' *heal*, *peal*, *peel*, *hill*, and *pill* are overwhelmingly classified with C2's *hill*; *heel* and *he'll* are predominantly classified as C2's *hill*, but with a substantial number of *heel* classifications (Table 12A). Utah *hail*, *hale*, *pail*, and *pale* are overwhelmingly classified with C2's *hill*; *hell* and *pell* are also generally classified with C2's *hell*, but with a substantial number of *hill* classifications (Table 12B). Finally, the Utah speakers' *pool* and *fool* are classified with C2's *hood* or *full*, while Utah *pull* and *full* are, for the most part, classified with C2's *whole* or *full* (Table 12C). In the Utah Two Speaker analyses, *heal*, *heel*, and *he'll* tend to be classified as *heal*, *heel*, or *he'll*; *peal* and *peel* tend to be classified as *hill*, *pill*, or *peel*; and *hill* and *pill* tend to be classified as *hill*, *peal*, or *peel* (Table 12A). Utah *hail*, *hale*, *pail*, and *pale* tend to be classified as *pale*, *pail*, or *hail*; *hell* and *pell* tend to be classified as *pell* or *hell* (Table 12B). And *pool* tends to be classified as *cook*; *fool* as *hoed*, *hole*, or *pole*; and *pull* and *full* tend to be classified as *pole*, *hole*, or *whole* (Table 12C).

Direct comparison of the Single Speaker analyses with both sets of Two Speaker analyses is revealing. Such a comparison involves not only the patterns of distinctiveness but also the specific confusions observed. *Hail*, *hale*, *pail*, and *pale* are classified differently from *hell* and *pell* in all three sets of analyses. In the Single Speaker analyses, each word tends to be classified correctly (Appendix B). In the Utah Two Speaker analyses, the words are not necessarily classified correctly, but most incorrect classifications involve the same VC Rhyme. But in the Connecticut Two Speaker analyses, *hail*, *hale*, *pail*, and *pale* differ from *hell* and *pell* primarily in the likelihood that they will be classified as *hill*; the /eɫ/ words are classified as *hill* more often than the /ɛɫ/ words are.

Table 13. *Comparison of Single Speaker (SS) and Two Speaker VC Rhyme discriminant analyses' classifications of /l/-final words. Only classifications assigned to 30% or more of the tokens with a particular Rhyme are listed. Full data for the Single Speaker analyses appears in Appendices D and E, and for the Two Speaker analyses in Appendices F and G.*

	/uɫ/	/ʊɫ/	/ʌɫ/	/oɫ/	/ɑɫ/	/ɔɫ/	/iɫ/	/ɪɫ/	/eɫ/	/ɛɫ/	/æɫ/
U6 SS	uɫ .94	uɫ .44 oɫ .44	oɫ .56	oɫ .45 uɫ .35	ɑɫ .38	ɔɫ .72	iɫ .80	ɪɫ .81	eɫ .91	ɛɫ .81	æɫ .69
< U8	ʊk .50	uɫ .88	oɫ .69	oɫ .88	æɫ .33	æɫ .34	eɫ .55	iɫ .50	eɫ .81	ɛɫ .63	æɫ .75
< U9	ʊk .50 ʊd .44	oɫ .63	oɫ .44 ʌɫ .38	oɫ .63 uɫ .31	ɑɫ .46	ɔɫ .44	ɪɫ .65 iɫ .33	iɫ .81	eɫ .91	ɛɫ .81	æɫ .50
< C2	ʊd .50	oɫ .69	oɫ .50	oɫ .75	†	ɑɫ .50	ɪɫ 1.00	ɪɫ 1.00	ɪɫ .63	ɪɫ .44 ɛɫ .38	æɫ .75
U8 SS	uɫ .62 ʊɫ .38	uɫ .31 ʊɫ .69	ʌɫ .75	oɫ .81	ɑɫ .33	ɔɫ .94	iɫ .60 ɪɫ .40	iɫ .38 ɪɫ .62	eɫ .65 ɛɫ .35	ɛɫ .69	æɫ .81
< U6	uɫ .62 ʊɫ .38	ʌɫ .81	ɑɫ .38	ʌɫ .66 ʊɫ .34	ɑɫ .50	ɑɫ .59 ɔɫ .31	ɪɫ .68	ɪɫ .81	iɫ .61	iɫ .44 ɛɫ .56	æɫ .75
< U9	uɫ .50	uɫ .94	ʌɫ .94	ʊɫ .56	ɑɫ .52	ɑɫ .53	iɫ .90	iɫ .69 ɪɫ .31	eɫ .55	ɛɫ .63	æɫ .63
< C2	ʊd .75	ʊɫ .50 ʊd .44	ʌɫ .63	oɫ .53	ʌɫ .48	ʌɫ .63	ɪɫ 1.00	ɪɫ .88	ɪɫ .74	ɛɫ .69	æɫ .81
U9 SS	uɫ .87	uɫ .38	ʌɫ .81	oɫ .56	ɑɫ .50 ɔɫ .50	ɑɫ .30 ɔɫ .66	iɫ .98	ɪɫ 1.00	eɫ .93	ɛɫ .94	æɫ .93
< U6	ʌɫ .69	oɫ .31	ʌɫ .38	oɫ .38	ɑɫ .83	ɑɫ .81	iɫ .70	iɫ .94	eɫ .53 iɫ .43	ɛɫ .75	ɛɫ .53
< U8	uɫ .69	oɫ .63	op .50	op .63	ɔɫ .54	ɔɫ .81	iɫ .48 ɪɫ .43	eɫ .56 ɛɫ .38	eɫ .87	ɛɫ .75	æɫ 1.00
< C2	uɫ .38 oɫ .31	uɫ .56	oɫ .44 ʊɫ .31	oɫ .59	ʌɫ .63	ʌɫ .67	ɪɫ 1.00	iɫ 1.00	ɪɫ .67 eɫ .30	ɛɫ .69	æɫ .93

†No one classification was assigned to at least 30% of the tokens, and six different VC classifications (with four different nuclei) were assigned to 10–24% of the tokens each.

Likewise, the Single Speaker analyses tend to classify *heal*, *heel*, *he'll*, *peal*, and *peel* and *hill* and *pill* as having the correct VC Rhyme (Appendix C). In both sets of Two Speaker analyses, *peal* and *peel* are classified with *hill* and *pill*, and this tendency is stronger in the Connecticut Two Speaker analyses. Finally, in the Single Speaker analyses, *fool* and *pool* are generally classified as having the correct VC (Appendix A), while *full* and *pull* are often classified as *hull*, *cull*, or a word containing /ol/. In the Utah Two Speaker analyses, *pool* tends to be classified as *cook*, *hull*, *pool*, or *full*; *fool* as *hoed*, *hole*, *pole*, or *full*; and *full* and *pull* as *hull* or a word containing /ol/. In the Connecticut Two Speaker analyses, *pool* and *fool* are classified as *hood*, and *full* and *pull* as *whole* or *full*.

The Two Speaker analyses are always less accurate than the Single Speaker analyses. However, direct comparison of the Utah and Connecticut Two Speaker analyses is not as straightforward. For the high and mid front vowels, the Connecticut analyses are worse than the Utah analyses. However, the Connecticut analyses are somewhat better than the Utah analyses for the high back vowels.

This picture is confirmed by the Two Speaker VC Rhyme analyses, as compared with the corresponding Single Speaker analyses for U6, U8, and U9. These comparisons are summarized in Table 13, which contains classifications received by 30% or more of the tokens with a particular Rhyme in the two sets of analyses;[22] the location of selected Utah VC Rhymes in C2's derived vowel space is illustrated in Figure 6. With regard to the contrasts of interest, the distinctions are always strongest in the Single Speaker analyses, as would be expected. Overall, /ul/-/ʊl/ are more distinct in the Utah Two Speaker analyses than in Connecticut Two Speaker analyses. For /el/ -/ɛl/, the Connecticut Two Speaker analyses outperform the Utah analyses for U8, but the Utah analyses better preserve the contrast for U6 and U9; for U6, both /el/ and /ɛl/ tend to be classified as C2's /ɪl/, although the tendency is stronger for /el/. Finally, Utah /il/ and /ɪl/, clearly distinct in the Single Speaker analyses, are not classified differently in the Two Speaker analyses, despite different mean coordinates for d1 and d2; while differences between /il/ and /ɪl/ are evident in the Utah Two Speaker analyses, their classifications are substantially less accurate than those in the Single Speaker analyses.

The basis for these classifications is evident in the canonical variable values plotted in Figure 6. In the three Utah panels of this figure, raw values for each speaker's vowels have been converted to values for the canonical variables based on C2's vowel space alone. Thus, in contrast with Figure 5, a single coordinate space is represented in all four panels of Figure 6. For all three Utah speakers, the front vowel combinations /il/, /ɪl/, /el/, when classified according to C2's vowels, are closer to her /ɪl/ than to her /il/ or /el/, and are so classified in the Two Speaker analyses. In interpreting Figure 6, it is important to bear in mind that only the first two canonical variables are plotted. So, even though U8's /il/ appears to be closer to C2's /el/ than /ɪl/ in Figure 6, the actual distance may be much greater on axes defined by higher order canonical variables.

4 DISCUSSION

4.1 Univariate vs. Multivariate Approaches

It is important to note that the picture of how contrasts are maintained derived from multivariate discriminant analysis—in Utah as in Connecticut—differs from the picture derived from consideration of the pairwise comparisons of vowel pairs along the raw acoustic dimensions in the univariate analyses summarized in Table 3. These differences are summarized in Table 14. In particular, at the third measurement point, none of C2's tense-lax vowel pairs before /l/ differs significantly in any parameter but f_0 and VQI. Yet it is F1 measured late in the vowel and not f_0 or VQI measured late in the vowel that makes the larger contribution to derivation of the canonical variables and to the discriminant analyses' distinction between the cognate tense-lax vowel pairs. Likewise, C4's vowel pairs can be distinguished at the third measurement point only in F2 (for /el/-/ɛl/). Nonetheless, F1 at the third measurement point contributes to the overall maintenance of the contrasts, as reflected in the discriminant analysis. Similarly, for all of the Utah subjects except U5, there are at least two parameters that contribute to the derivation of one or both canonical variables while themselves distinguishing at most one cognate tense-lax pair. The multivariate approach, we feel, corresponds better to ordinary language use, in which individuals produce and perceive complex sounds that vary simultaneously along a number of dimensions rather than producing or perceiving any one dimension independently of the others.

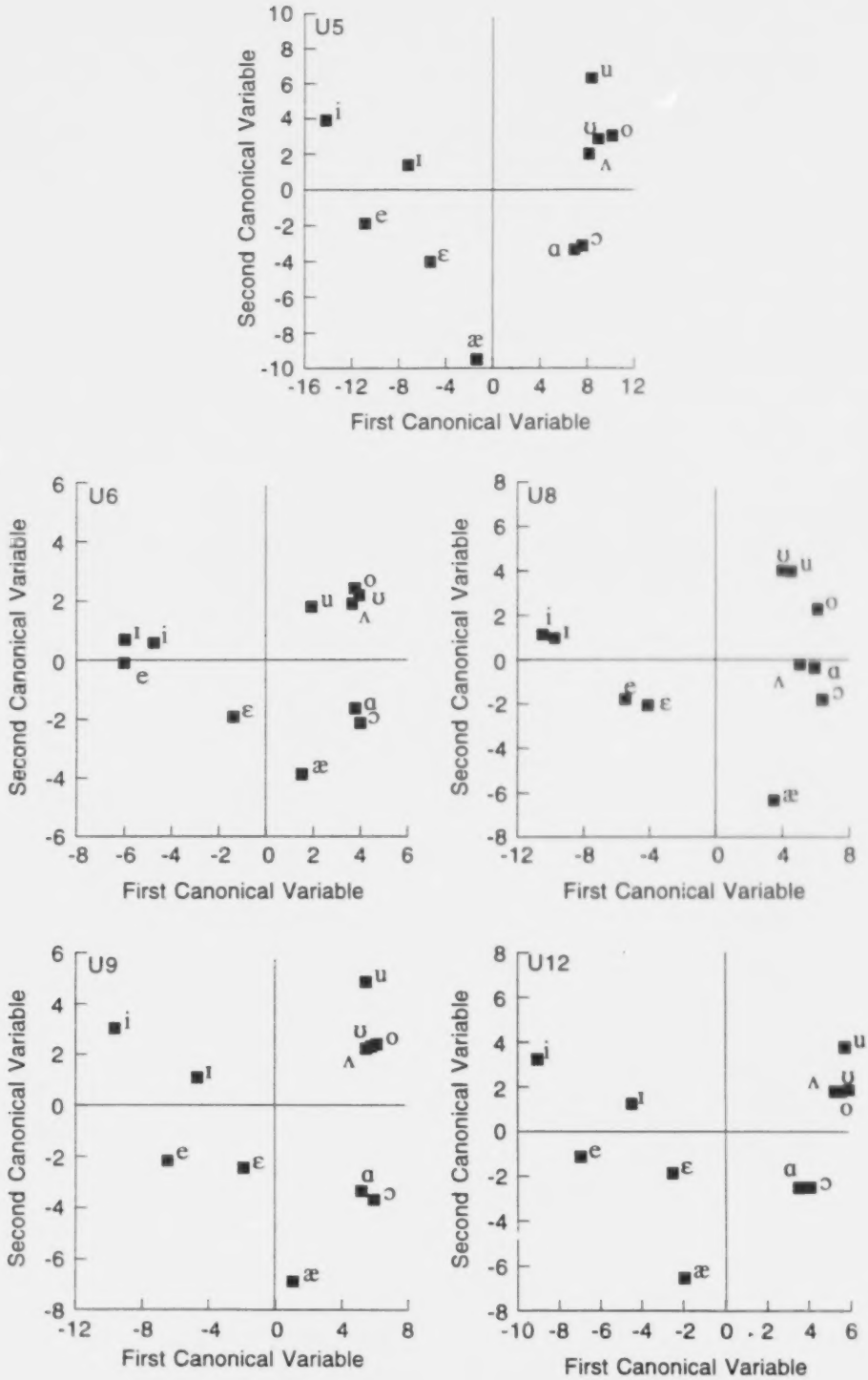


Figure 6. **First two canonical variables from Two Speaker general VC Rhyme discriminant analyses for female subjects C2, U6, U8, and U9. Solid squares represent mean values for the individual speakers' VCs, and, to simplify inter-speaker comparison, crosses on the Utah speakers' figures represent C2's /il op ol ud/.**

Table 14. *Variables used in Single Speaker discriminant functions (/l/ words only) that, based on the Post-hoc tests summarized in Table 3, serve to distinguish one or no cognate tense-lax pairs before /l/.*

Subject	Parameter	Location	Pairs Distinguished
C2	F1	Late	0
C4	F1	Late	0
	VQI	Middle	0
U6	VQI	Early	0
		Middle	0
	F1	Early	1
		Middle	1
U8	F2	Early	0
		Middle	1
	F1	Early	0
		Middle	0
		Late	0
	VQI	Early	1
U9	VQI	Early	0
		Middle	0
U12	F1	Middle	1
	f0	Early	0
	VQI	Early	0

One additional surprising result of our multivariate approach is worth noting: f_0 makes almost no contribution to any of the discriminant functions, despite the well-known intrinsic f_0 differences among vowels (e.g., Lehiste & Peterson, 1961). That is, despite the fact that vowels differ systematically in fundamental frequency, these differences do not contribute to maintaining vowel contrasts. This is true of the VC Rhyme /l/-only analyses discussed here as well as of the general VC Rhyme analyses. Despite the ubiquity of intrinsic f_0 variation, in the literature as in the present corpus (Section 3.1 above), there is extensive inter-speaker variation in the relative ranking of vowels in f_0. In particular, speakers tend to differ as to whether /i/, /ɪ/, /u/, or /ʊ/ has the highest f_0, and, as a result, f_0 is unlikely to provide a consistent cue to vowel identity in English.

An additional feature of the single-speaker VC Rhyme analyses (both general and /l/ only) that is worthy of note is the essential congruence of the canonical variables isolated. For all speakers, Utah and Connecticut, the first canonical variable defines a vector ranging from high, front vowels to low, back vowels. For four out of the seven speakers, the second canonical variable defines a vector from low, front to high, back vowels; for two additional speakers, d2 is a vector from low to high vowels; and for the last speaker, it is a vector from low, fronting to high, backing vowels. Where speakers differ is in which end of a canonical variable vector, if any, is associated with breathy phonation and which with creaky phonation. (They also differ as to whether f_0 contributes to derivation of the canonical variables and in the exact weight given to the input variables in derivation of the canonical variables.)

In order to understand the congruence of the VC Rhyme discriminant analyses, it is necessary to review how discriminant analysis derives the canonical variables. Given the 12 input variables in these analyses, it constructs a 12-dimensional coordinate space within which all of the tokens can be situated. It then finds the single longest vector through the point cloud(s) representing the tokens. This is the first canonical variable. In effect, it is the vector through the coordinate space along which the largest spread in the data is observed. The second canonical variable is the vector along which the maximum remaining range of variation is observed, and so on.

The traditional vowel quadrilateral, illustrated in Figure 7, can be defined in terms of subjective tongue position or in terms of the first two formants. Because it is a quadrilateral and not a square, its two diagonals are not equal in length. Assuming that a speaker's vowels are distributed equally throughout the vowel space, the longest vector through the space is likely to correspond to the high, front to low, back dimension. Given that all of the data for each speaker were standardized, the heavy contribution of F2 to the first canonical variable is unrelated to the fact that F2 typically ranges over c. 2500 Hz, as compared with c. 700 Hz for F1; because of the larger standard deviation for F2, *z*-scores for F1 and F2 are comparably distributed.

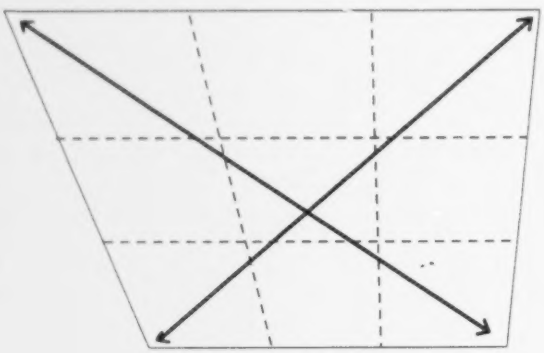


Figure 7. **Stylized vowel trapezoid, showing the relationship between the cells defined by the traditional division into high, mid, and low vowels, and front, central, and back vowels and the two diagonals.**

Assuming that phonation differences are indeed relevant in describing English vowels, we would expect breathy phonation to be associated with the high, front end of d1 and with the high, back end of d2. In both cases, the association of breathiness with one end of the vector or the other would maximize variation along that vector. However, since the association of breathiness with particular supra-laryngeal configurations is based on other factors than the geometry of the vowel trapezoid, it is less rigid. Thus, differences among speakers can be expected. In the present sample, it appears that association of creakiness rather than breathiness with the high, front end of d1 may be indicative of change in progress. U6, one of the speakers with this pattern (see Table 10), has a vowel space in which /ɪl/ has substantially higher F2 than does /il/ (see Figure 4, p. 87). Thus, the different role of breathiness in derivation of her canonical variables reflects this reversal of nuclei in formant space rather than changes in the relative breathiness of /ɪl/ and /il/.

4.2 Patterns of Change

Our results to this point give rise to a question that can only be answered speculatively: How did the patterns that we have described come about? We know that a series of changes began at least 25–30 years ago, with /ul/ and /ʊl/. Labov, Yaeger, and Steiner (1972) report a similar phenomenon in Albuquerque, NM, and in Salt Lake City, and confusion between /ul/ and /ʊl/ in transcription exercises by University of Utah students began c. 25 yrs. ago (Wick Miller, personal communication). Our data reflect two expansions of that original approximation of /ul/ and /ʊl/. The first is the expansion to the other tense-lax vowel pairs before /l/, and the second is the development of implementations of the tense-lax contrast before /l/ that have not been reported for other dialects of English. This second expansion may be related to the approximation of /ʌl/, /ʊl/, and /ol/ for some of our speakers. The first expansion does not presuppose maintenance of the tense-lax contrast despite the close approximation of the contrasting elements, but the second expansion does. There are two possible bases for maintenance of the contrast. Either it is a reimportation from other dialects of English, or there is something in Utah speech maintaining the contrast, despite the approximation. It is worth noting that, given the original close approximation of /ul/-/ʊl/, /il/-/ɪl/, and /el/-/ɛl/, morphological support for associating /ul/, /il/, /el/ with /u/, /i/, /e/, comes from the pronominal system, in which *you'll*, *he'll*, *she'll*, *we'll*, *they'll* are derivational variants of *you*, *he*, *she*, *we*, *they*.[23] In fact, it is conceivable that vowel quality variations in *-ll* contractions, as observed, for example, in subject C4 in the present sample, provide the original basis for the approximation of tense and lax vowels before /l/, as well as for the maintenance of the contrast.

We have previously argued (Di Paolo & Faber, 1990), and continue to believe, that spectral slope differences contribute to maintenance of the contrast between tense and lax vowels before /l/ in Utah. Faber (1992) suggests a basis for reinterpretation of the contrast between tense and lax vowels as one based on phonation. As noted above, in vowels with low F1 like /i/ and /u/, the fundamental is likely to fall within the F1 resonance, and thus these vowels will be characterized by spectrally prominent f_0. Breathy phonation likewise increases the spectral prominence of f_0. That is, both raising the tongue body sufficiently to produce a very low F1 and producing a vowel with breathy phonation have similar effects on spectral profile. Given such a contrast, language learners might impute it to tongue position differences, to phonation differences, or to both, covarying.[24] Indeed, some subjects in the perception study described in Di Paolo & Faber (1990) responded differentially to vowel nuclei differing only in VQI, while others did not, suggesting that to some speakers, but not others, phonation differences play a crucial role in the tense-lax contrast before /l/. U8 may be one of those speakers. In her system, tense and lax vowels before /l/ are only minimally distinguished, if at all, by a subtle combination of formant and phonation differences.

There are three possible outgrowths of such a vowel system, two of which are exhibited in the present corpus. The first possibility is that, under the influence of other dialects of English, formant differences will once again become widespread, so that Utah English will cease to differ from other dialects in this respect. This pattern is manifested by the two male subjects, U5 and U12, both of whom were, subjectively, extremely careful speakers. U5 is also the only Utah subject in the present sample with substantial exposure to other varieties of English, both as a two-year resident of Toronto and as a teacher of English as a Second Language in Taiwan for three years. Secondly, it is possible that at some point a true merger will take place, making whatever observable differences there still might be between *heel* and *hill* qualitatively as well as quantitatively comparable to those between *heel* and *heal*. This is the pattern that U8 appears at first blush to have, although

analysis in greater depth reveals, as we have shown, that she does indeed maintain the contrasts. Finally, it is possible that additional acoustic differences will be associated with the tense-lax contrast before /l/, so that the contrast ultimately will be phonetically different in the near merger area than in other parts of the English speaking world. Two related phenomena suggest to us that the latter is the case. Both are evident in U6's data (Figure 4). The first phenomenon suggesting an on-going reimplementation of the tense-lax contrast is the apparent reversal of /il/ and /ɪl/ in F1, F2, or both. The second is the increase of F2 for /ul/, substantially lagging behind this change in other phonological contexts; U9 has a similar, albeit smaller magnitude, increase in F2 for /ul/. In both cases, the original lax vowel is now more peripheral in the vowel space than is the original tense vowel.

Another phenomenon regarding vowels before /l/ that we have already alluded to in our discussion is the approximation of /ʌl/, /ol/, and /ʊl/. Aside from the general United States pattern in which all three nuclei are clearly distinct, and an alternative pattern in which /ʌl/ and /ʊl/ are at best marginally distinct from each other but both are clearly distinct from /ol/ exhibited by C1 (Figure 3), we observe all logically possible combinations:[25] for two of the speakers described in Di Paolo & Faber (1990), /ʌl/ and /ol/ appear to overlap but are clearly distinct from /ʊl/; for four of the speakers, /ʊl/ and /ol/ overlap but are distinct from /ʌl/; and for an additional six subjects, all three nuclei appear to overlap.[26] In the present sample, C2, C4, and U8 have the general pattern, C1 and U5 have the alternative pattern, U9 has /ʊl/-/ol/ distinct from /ʌl/, U6 has closely approximated /ʊl/, /ʌl/, and /ol/, and U12 apparently has three overlapping nuclei.

The speakers described in Di Paolo and Faber (1990) came from two socioeconomically and geographically distinct regions in the Salt Lake Valley, the predominantly white collar Eastside and the predominantly blue collar Westside. Of the three speakers from the Salt Lake Valley in the present sample, U6 and U9 are from the Eastside, and U12 is from the Westside. In the earlier study, we found that the tendency for formant reversals in cognate tense-lax pairs before /l/ was predominantly a Westside tendency. In terms of the current study, all of the speakers showing full overlap of /ʌl/, /ol/, and /ʊl/, six from the Di Paolo & Faber (1990) sample as well as U12, are from the Westside, while the other four patterns are evenly divided between the Eastside and the Westside. U5 and U8 of course are not from the Salt Lake Valley, and, like the Eastside speakers, are not participating in the Westside pattern.

The structural relationship between the phenomena under discussion is clear: /ʊl/ can, in principle, merge with /ul/, or it can merge with /ʌl/ and/or /ol/. However, on purely functional grounds, a four-way merger seems less likely, and, indeed, does not occur in any of our corpora. At present, any scenario relating the two systematic changes must remain speculative, due to insufficient data regarding /ʌl/ and /ol/. It may be that the three-way approximation of /ʌl/, /ʊl/, and /ol/ represents the most recent stage in a series of diachronic developments in which /ul/-/ʊl/ approximated, their approximation was generalized to /il/-/ɪl/ and /el/-/ɛl/, and then /ul/-/ʊl/ (and possibly /il/-/ɪl/ and /el/-/ɛl/) diverged again. Alternatively, both the /ul/-/ʊl/ and /ʌl/-/ʊl/-/ol/ approximations may be competing resolutions to a perceived instability of /ʊl/. Full evaluation of these competing possibilities must await fuller investigation of vowel systems in which comparable approximations of vowels before /l/ have been reported. Only such an investigation will enable a determination of whether the precise phenomena observed in Utah are unique to Utah or whether they are characteristic responses to perceived dialect conflict in cases of urbanization in which dialects with southern admixture come into closer contact with dialects without such admixture. In any case, these approximations of vowels before /l/ are similar to, and perhaps related to, the general approximation of /ɯ/,[27] /o/ and /ʌ/ before /r/, which also appears to be moving toward merger. The vowels /o/, /ɔ/, and /ɑ/ before /r/ have also been participating in various reported approximations and mergers in the Western United States (e.g., Labov, Yaeger, & Steiner, 1972; Yaeger, 1974; Stanley, 1936; Norman, 1971).

5 CONCLUSION

In summary, we have found that discriminant analysis is a worthwhile technique with which to study potential linguistic contrast. In particular, our analyses show that vowel pairs that are not distinct along any one measurable dimension—F1, F2, f_0 or spectral slope—may be distinct when all dimensions are considered simultaneously. This result underlines the importance of describing vowels in terms of as many dimensions as possible, not merely those dimensions that, like the first and second formant frequencies, are *a priori* assumed to distinguish among vowels. We

have not addressed in this paper whether or not Utah listeners actively discriminate the elements of these near-merged contrasts. (See Di Paolo & Faber [1990] and Faber, Best, & Di Paolo [1993, 1994] for evidence that at least some Utahns do perceive the difference.) What we have shown is that by and large Utah speech contains sufficient information to enable Utah listeners to perceive the contrasts. Our emphasis has been on potential distinctiveness, which, after all, is logically prior to actual discrimination by listeners. This emphasis differs from that of Labov, et al. (1991) who suggest that listeners may not make use of minimal distinctions in actual language use.

Our goal in this paper has been to describe vowel contrast in terms of a multi-dimensional acoustic space, and, indirectly, the articulatory space therein reflected. The statistical technique that we used, discriminant analysis, distinguishes among pairs of words that are acoustically very similar (and, perhaps, indistinguishable to outsiders). These small differences in vowel location in a multi-dimensional acoustic space (reflecting small articulatory differences) suffice to preserve phonological contrasts that may not be evident to the naked introspective ear, and thus provide the basis for possible future enhancement of the contrasts.

Given that there are speakers with near mergers of /il/-/ɪl/, of /el/-/ɛl/, and of /ul/-/ʊl/, if these facts were presented in purely phonological terms, the latter developments would appear to reflect reversal of an absolute merger, of the sort that has been appealed to as an account of the *meat-mate* facts in the past (Labov, 1974; Milroy & Harris, 1980; Faber, Di Paolo, & Best, 1994), and as an account of the apparent loss of a rule of final devoicing in most (if not all) dialects of Yiddish (King, 1980; Faber, Di Paolo, & Best, ms). But this is not what is happening in Utah. By providing a detailed mechanism by which contrast can be preserved and enhanced in the case of one near merger, we hope to have cast further suspicion on the notion of "reversal" of merger.

REFERENCES

Allen, H. B. (1973–76). *The linguistic atlas of the Upper Midwest*. 3 Vols. Minneapolis: University of Minnesota Press.

Bailey, G., Winkle, T., Tillery, J., & Sand, L. (1991). The apparent time construct. *Language Variation and Change, 3*, 241–264.

Bond, Z. S. (1973). The perception of sub-phonemic differences. *Language and Speech, 16*, 351–355.

Colton, R. H., & Conture, E. G. (1990). Problems and pitfalls of electroglottography. *Journal of Voice, 4*, 10–24.

Di Paolo, M. (1988). Pronunciation and categorization in sound change. In K. Ferrara, B. Brown, K. Walters, & J. Baugh (Eds.), *Linguistic change & contact: Proceedings of the Sixteenth Annual Conference on New Ways of Analyzing Variation in Language* Austin: *Texas Linguistic Forum, 30*, 84–92.

Di Paolo, M. (1992a). Hypercorrection in response to the apparent merger of (ɑ) and (ɔ) in Utah English. *Language & Communication, 12*, 267–292.

Di Paolo, M. (1992b). Evidence for the instability of a low back vowel 'merger.' Presented at the Twenty-First Annual Conference on New Ways of Analyzing Variation in Language.

Di Paolo, M, & Faber, A. (1990). Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language Variation and Change, 2*, 155–204.

Dixon, W. J., (Ed.), (1988). *BMDP statistical software manual*. Berkeley: University of California Press.

Faber, A. (1986). On the actuation of sound change: A Semitic case study. *Diachronica, 3*, 163–184.

Faber, A. (1992). Articulatory variability, categorical perception, and the inevitability of sound change. In G. Davis & G. Iverson (Eds.), *Explanation in historical linguistics* (pp. 58–75). Amsterdam: John Benjamins.

Faber, A, Best, C. T., & Di Paolo, M. (1993). Dialect differences in vowel perception. Presented at the Fall meeting of the Acoustical Society of America.

Faber, A, Best, C. T., & Di Paolo, M. (1994). Dialect differences in production and perception. Presented at the Fall meeting of the Acoustical Society of America.

Faber, A, Di Paolo, M., & Best, C. T. (1994). The peripatetic history of Middle English *ɛ̄. Paper presented at the Annual Meeting of the Linguistic Society of America.

Faber, A, Di Paolo, M., & Best, C. T. (ms.). Perceiving the unperceivable: The acquisition of near merged contrasts.

Fant, C. G. (1956). On the predictability of formant levels and spectrum envelopes from formant frequencies. In M. Halle *et al* (Eds.), *For Roman Jakobsen* (pp. 109–119). The Hague: Mouton.

Fowler, C. A. (1994). Invariants, specifiers, cues: An investigation of locus equations as information about place of articulation. *Perception & Psychophysics, 55*, 597–610.

Harrington, J., & Cassidy, S. (1994). Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs. *Language & Speech, 37*, 357–373.

Harris, J. (1985). *Phonological variation and change*. Cambridge: Cambridge University Press.

Hartman, J. W. (1984). Some possible trends in the pronunciation of young Americans (maybe). *American Speech, 59*, 218–225.

Henton, C. (1992). Acoustic variability in the vowels of female and male speakers. Paper presented at the Spring meeting of the Acoustical Society of America.

Javkin, H., Antoñanzas-Barroso, N., & Maddieson, I. (1987). Digital inverse filtering for linguistic research. *Journal of Speech and Hearing Research, 30*, 122–129.

Johnson, K. (1989). On the perceptual representation of vowel categories. *Indiana University Research on Speech Perception, Progress Report, 15*, 343–358.

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel perception. *Journal of the Acoustical Society of America, 94*, 701–714.

King, R. D. (1980). The history of final devoicing in Yiddish. In M. I. Herzog, B. Kirschenblatt-Gimblett, D. Miron, & R. Wisse (Eds.), *The field of Yiddish IV* (pp. 371–430). Philadelphia: Institute for the Study of Human Issues.

Klecka, W. R. (1980). *Discriminant analysis* (Sage University Paper Series: Quantitative Applications in the Social Sciences, 19). Newbury Park, CA: Sage Publications.

Kurath, H., & McDavid, R. I. (1961). *The pronunciation of English in the Atlantic states*. Ann Arbor: University of Michigan Press.

Labov, W. (1974). On the use of the present to explain the past. In *Proceedings of the eleventh International Congress of Linguists* (pp. 825–851). Vol. 2. Bologna: Società editrice Il Mulino.

Labov, W. (1991). The three dialects of English. In P. Eckert (Ed.), *New ways of analyzing sound change* (pp. 144). Orlando: Academic Press.

Labov, W. (1994). *Principles of linguistic change: Internal factors*. London: Blackwells.

Labov, W., Yaeger, M., & Steiner, R. C. (1972). *A Quantitative study of sound change in progress* (Report on NSF contract GS-3287). Philadelphia: US Regional Survey.

Labov, W., Karen, M., & Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change, 3*, 33–74.

Ladefoged, P. (1983). The linguistic use of different phonation types. In D. H. Bless & J. A. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 351–360). San Diego: College Hill.

Lane=Kurath, H., (Ed.) (1939–43). *Linguistic atlas of New England*. Providence: Brown University.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America, 33*, 454–460.

Lofqvist, A. (1991). Inverse filtering as a tool in voice research and therapy. *Scandinavian Journal of Logopedics and Phoniatrics, 16*, 8–16.

Milroy, J., & Harris, J. (1980). When is a merger not a merger? The MEAT/MATE problem in a present-day English vernacular. *English World Wide, 1*, 199–210.

Norman, A. (1971). A Southeast Texas dialect study. In H. B. Allen & G. N. Underwood (Eds.), *Readings in American dialectology* (pp. 135–151). New York: Appleton-Century-Crofts.

Nunberg, G. (1980). A falsely reported merger in eighteenth century English: A study in diachronic variation. In W. Labov (Ed.), *Locating language in time and space* (pp. 221–250). New York: Academic Press.

Orton, H., et al. (1969–). *Survey of English dialects*. Leeds: E. J. Arnold.

Pederson, L. A., McDaniel, S. L.; Bailey, G., & Bassett, M.. (1986). *Linguistic atlas of the Gulf States. Vol. 1: Handbook for the linguistic atlas of the Gulf States*. Athens, GA: University of Georgia Press.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175–184.

Port, R., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics, 17*, 257–282.

Stanley, O. (1936). The speech of East Texas. *American Speech, 11*, 3–36, 145–166, 232–251, 327–355.

Stevens, K. N. (1988). Modes of vocal fold vibration based on a two-section model. In O. Fujimura (Ed.), *Vocal fold physiology: Voice production, mechanisms and functions* (pp. 357–370). New York: Raven.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3–45.

Sussman, H. M. (1991). The representation of stop consonants in three dimensional acoustic space. *Phonetica, 48*, 18–31.

Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America, 90*, 1309–1339.

Wells, John C. (1982). *Accents of English*. (3 Vols.). Cambridge: Cambridge University Press.

Whalen, D. H. (1991). Infrequent words are longer in duration than frequent words. *Journal of the Acoustical Society of America, 90*, 2311 (A)

Yaeger :, M. (1974). Speaking style: Some phonetic realizations and their significance. *Pennsylvania Working Papers on Linguistic Change and Variation, 1*, 1–60.

FOOTNOTES

Language Variation & Change, in press (1995).

†University of Utah.

[1]Scattered survey work, informal observations and anecdotal evidence in the Southern and Western United States report similar phenomena of unknown geographical extent. To our knowledge, the first examination of this sound change appeared in Labov, Yaeger, and Steiner (1972). Subsequently, there have been a number of other reports. For example, Hartman (1984) reports near mergers before /l/ from California to Kansas. Bailey, Winkle, Tillery, and Sand (1991) observe mergers of tense and lax vowels before /l/ in younger urban Texas speakers. Tom Clark (Personal Communication) has also reported confusions between /eɪ/ and /ɛɪ/ in, e.g., *bail-bell* in the Las Vegas, NV area. Orthographic confusion between /eɪ/-/ɛɪ/ is common in the Salt Lake Valley (e.g., *Bake Sell*), and we have also received second and third-hand reports of similar confusions in Taos, NM (over a plate of free samples in a supermarket *Fill free to taste*) and Florida (from an undergraduate essay *She feels the air with joy and happiness*); we thank Nicole Kikowski and David Johns for these examples. The eye-dialect form *rilly*, widely used to characterize so-called Valley Girl speech, may also reflect a similar phenomenon. Outside the US, merger of /i-ɪ/ and /u-ʊ/ before /l/ is reported in London vernacular (Wells, 1982). In addition, the *Survey of English Dialects* (Orton, et al., 1969-) includes scattered instances that *may* reflect merger or near merger (e.g., *nail* [nɛl], *meal* [mɪl]) in the Midlands and the North Counties. We do not know how general these phenomena are or to what extent they are acoustically and perceptually comparable with the Utah phenomenon to which we have devoted our attention. There has been relatively little systematic dialect survey west of the Mississippi comparable to that done east of the Mississippi (e.g., Kurath & McDavid, 1961; Allen, 1973–76; Pederson, McDaniel, Bailey, & Bassett, 1986) so the apparent merger of cognate tense and lax vowels before /l/ might be much more widespread in the Western United States than this scattered evidence would suggest.

[2]The categorization task reported in Di Paolo (1988) and Di Paolo & Faber (1990) likewise represents a relatively formal style.

[3]Cf. Bond (1973) on duration differences between homophones based on morphological structure (e.g. *laps* vs. *lapse*) and Whalen (1991) on duration differences based on lexical frequency (e.g., *ewe* vs. *you*).

[4]Of course, *heel* and *heal* originally were not homophones. Cf. the discussion in the text of *meet* and *meat*.

[5]Our original intention had been to split the analysis so that approximately half the data would be analyzed at each site. However, it proved much less time consuming to analyze data on the Haskins VAX cluster than on the Macintosh available at the University of Utah.

[6]Formant measurements for the subjects analyzed at Haskins were made via the ILS LPC-based root solving algorithm (RSO); for the subject analyzed at the University of Utah, formant measurements were made from narrow-band DFT spectra produced by Signalyze. F_0 for the Haskins subjects was arrived at by a combination of the ILS subprogram API and hand-measuring of individual pitch periods; for the Utah subject, f_0 was determined using the Signalyze cepstral

analysis routines. For all subjects, VQI was computed from narrow band DFT spectral cross-sections, produced either by Signalyze or by the Haskins Laboratories program HADES, by subtracting the amplitude of the strongest harmonic in F1 from the amplitude of the first harmonic. While it would have been desirable to include duration measurements, as well, this was not possible. Vowel duration could not be measured because of the difficulty of reliably segmenting between a vowel and a following /l/. Combined duration of the vowel and the following consonant could not be measured for all tokens, as some final stops for most subjects were not released, providing no release burst on the basis of which to make the measurements. For subject C1, only formant measurements are available. Background noise in the recording (the first done in a new set-up) made f_0 measurement difficult and made amplitude measurements dangerous.

[7]The data were standardized with the z-transform in order to simplify interpretation of the Two Speaker analyses, in which each Utah female speakers' vowels were classified in terms of C2's vowel space and in terms of those of the other two Utah females. Even though the Two Speaker analyses only involved the female speakers, the male speakers' data were standardized as well. We computed separate z-scores for each of the four measured parameters for each speaker. For example, we computed the mean and standard deviation for U6's F1 across all three measurement locations and all 440 tokens, and, on that basis, converted each F1 value to its corresponding z-score, and so forth for the other three parameters and for all speakers.

[8]Even though our purpose is to determine which dependent variables make substantive contributions to each discriminant function, the stepwise procedure that program 7M defaults to, which would ostensibly select all and only the variables that make such a contribution, is inappropriate. This is because the stepwise procedure provides non-unique solutions. That is, multiple runs using the same data may arrive at different discriminant functions (discriminant functions containing different variables), depending on the order in which variables are entered into the analysis. In the present series of analyses, the non-uniqueness problem was avoided by forcing all dependent variables into the discriminant function, and then eliminating from further consideration those variables that had a partial F value of less than 3.00 at the final step in the analysis.

[9]Discriminant analysis is one of a family of multivariate statistical techniques that can be used for assessing the reliability of an a priori division of a data set into subgroups, and for concentrating the variability in a multi-dimensional data set onto a smaller number of (derived) dimensions. Related techniques include factor analysis, cluster analysis, and principal components analysis. Our choice of discriminant analysis rather than one of these other techniques was based on its generality, in allowing simultaneous consideration of the group membership of individual tokens and of the dimension(s) along which these groups differ. While other techniques (e.g., principal components analysis: Harrington & Cassidy, 1994) may give comparable accuracy in distinguishing among groups in a data set, we are equally interested in the dimensions along which these groups differ.

[10]For example, given three input variables *length*, *width*, and *height*, the equation deriving a given canonical variable from the input variables will be of the form:

$$CV = (x * length) + (y * width) + (z * height) + k,$$

where k is a constant. The greater the coefficient x, y, or z, the greater the contribution of the variable it is associated with to discriminant function CV.

[11]This is Pattern 3 of Labov, Yaeger, & Steiner (1972), observed in British and Southern United States dialects. Labov (1991) notes an expansion of this 'Southern Shift' to other United States dialects.

[12]For discussion of maintenance of a minimal /ɑ/-/ɔ/ contrast despite the overt norm in which the contrast is absent, see DiPaolo (1992a, 1992b).

[13]In this class, /ʌ/ and /ʊ/ reflect different developments from Middle English *u. When followed by /l/, the common Northeastern US reflex is /ʌ/, except in a handful of words, typically described (e.g., Wells, 1982) in terms of their initial labial consonants. Thus, for most speakers, *full* and *cull* indeed have different vowel nuclei (/fʊl/ vs. /kʌl/). However, pockets of apparent lack of contrast, in which *bulge*, for example, is /bʊlj/ rather than /bʌlj/, have been observed in New England (Kurath & McDavid, 1961; LANE, map 362). Subject C1 is from Middletown, one of the areas in Connecticut where this pattern was observed in the 1930's. C1's overlap of /ɑl/ with /ʌl/-/ʊl/ may be a result of her substituting [ʌl] for /ɑl/ in the (to her) novel items *Sol*, *Col*, *pol*.

[14]Because only formant measurements were available for C1, no discriminant analyses were performed on her data.

[15]Some colleagues of ours from throughout the US have reported to us that they, too, tend to group *he'll* with *hill* rather than with *heel*, but we have no evidence for how widespread this pattern might be and what it correlates with.

[16]These significance levels should be interpreted as follows: Given that the 11 vowels before /l/ give rise to 55 pairwise comparisons, for an overall significance level of .05, the criterion for each individual comparison is .05/55 or .0009.

[17]These rankings represent *independent* contributions to the discriminant function. Given that F1 amplitude is dependent on F1 frequency (Fant, 1956), regardless of underlying laryngeal configuration, VQI values will tend to be correlated with F1 frequency. However, F1 frequency does not account for all of the variance in VQI. When VQI is included in derivation of a canonical variable, it is on the basis of the residual variance not resulting from variation in F1.

[18]The canonical variables in the more general VC Rhyme analyses serve to distinguish among VC rhymes on the basis of C as well as of V. In the present instance, our interest is in how vowels before /l/ differ, and it is easier to glean this information from the /l/-final analyses than from the general VC Rhyme analyses.

[19]It is essentially arbitrary which end of the vector is treated as negative and which as positive (Klecka, 1980); thus, signs have been changed on some coefficients and vectors to simplify inter-subject comparisons.

[20]Vowels with low frequency F1 will also have low L1 (Fant, 1956) and may also have high L0, if F1 is low enough that the fundamental will fall within its bandwidth. Thus, high vowels will naturally tend to have spectral cross-sections that are comparable to those of vowels (of any height) produced with breathy phonation. As a result, if VQI were varying only as a function of vowel height, high breathy vowels (those with low F1 and high VQI) and low creaky vowels (those with high F1 and low VQI) would be expected to occupy extremes of the vectors defined by the canonical variables whose derivation they contribute to. In such cases, F1 and VQI should have opposite signs. Of the 18 VQI-F1 pairings observable in Table 8 (two each for C2, C4, U5, U8, and U12, and four each for U6 and U9), this natural association only occurs four times, for C2's d2, U5's d1, and U8's d1 and d2. In three additional cases, the VQI coefficient is near 0, and hence irrelevant: VQI-early for U6's d2, VQI-mid for U9's d1, and VQI-early for U12's d2. The

F1-mid coefficient is near 0 for U6's d1, while the VQI-mid coefficient is negative. In the other 9 cases, at least one for all subjects but U8, low F1 and creaky VQI comparably increase or decrease the canonical variables whose derivation they participate in. This covariation necessarily reflects active control of phonation type independently of variation in tongue position, as reflected in F1 frequency.

[21]This analysis was restricted to the female speakers because of the possibility that female speakers in general have larger vowel spaces than males do, even after normalization designed to compensate for females' generally higher frequency resonances (Henton, 1992; Johnson, 1989). C2 was used as representative of Connecticut speakers, because of the restriction of C1's data to formant measurements alone.

[22]The full data for the Single Speaker analyses appear in Appendices D and E and for the Two Speaker analyses in Appendices F and G.

[23]We are extremely grateful to Cathi Best for this observation.

[24]This follows from the view of sound change outlined in Faber (1986, 1992).

[5]The general and the alternative patterns are the only two for which we have even anecdotal evidence outside our main study area.

[26]We cannot categorize our balanced word list speakers, BW and NM (described in Di Paolo & Faber, 1990), along this dimension, since the balanced word list contained no words with /ol/. Also, given that we have not studied the /ʌl/-/ol/-/ul/ problem in any great depth, we prefer not to talk about merger or lack of contrast; we merely note that for some speakers, some nuclei are closer together than they are for other speakers. In particular, the distinction between the /ul/-/ol/ overlapping and the /ʌl/-/ol/ overlapping groups may be artifactual, since there were very few data points for any one nucleus in the Di Paolo & Faber (1990) samples.

[27]The vowel /ur/ in this dialect, as in others, reflects earlier /ur/ and /ʊr/, just as /or/ in the perceived standard of the area reflects earlier /or/ and /ɔr/.

APPENDIX A

Classification of words with high back vowels before /l/ from Single Speaker target Word discriminant analyses. Figures represent the number of tokens (out of 8) for which each classification occurred. The panel for U8 appears also in Table 5.

	Class\ Target	fool	pool	full	pull	whole	hull / cull	poll / pole	other
C2	fool	6	2						
	pool	3	5						
	full			6	2				
	pull			1	7				
C4	fool	7	1						
	pool		7					1	
	full			3	3		1		1
	pull			3	5				
U5	fool	6	2						
	pool	5	3						
	full			7			1		
	pull			2	3	1	1	1	
U6	fool	7					1		
	pool		8						
	full			3	1	2		1	1
	pull			1		2	1	2	2
U8	fool	3	1	2	1			1	
	pool	2	3	2	1				
	full	1	1	3	2			1	
	pull	1	2	2	3				
U9	fool	5	1	2					
	pool		6			1		1	
	full	1	2	3				1	1
	pull				2		4	2	
U12	fool	5	2		1				
	pool	4	4						
	full			2			3	2	1
	pull			1	3		1	1	2

APPENDIX B

Classification of words with mid front vowels before /l/ from Single Speaker target Word discriminant analyses.

	Class / Target	hail	hale	pail	pale	hell	pell	hill / pill	other
C2	hail	5	3						
	hale	2	5	1					
	pail	1		2	5				
	pale		1	6	1				
	hell					4	4		
	pell					3	5		
C4	hail	2	4	1	1				
	hale	3	3	2					
	pail	2	1	4	1				
	pale	2	2	3	1				
	hell					5	2	1	
	pell						8		
U5	hail	1	1	4	2				
	hale	2	4	1	1				
	pail	2	2	2	2				
	pale	5	1		2				
	hell					5	3		
	pell					2	5		1
U6	hail	3	2	1	1			1	
	hale		3	3	2				
	pail	1	2	4	1				
	pale	3	1	1	1				2
	hell						3	3	2
	pell		1			2	5		
U8	hail	5			1	1	1		
	hale	1		2	2	2			
	pail	2	2	1	1	2			
	pale	2	3	2		1			
	hell	1	2	1		1	2	1	
	pell		1		1		6		
U9	hail	5	3						
	hale	2	4	1	1				
	pail	2	2	2	1			1	
	pale†			2	4				
	hell				1	6	1		
	pell					1	6		1
U12	hail	4	1	1	1				1
	hale	1	1	2	3				1
	pail	2	1	1	4				
	pale		3	4	1				
	hell†					3	3		
	pell					3	5		

† indicates fewer than 8 tokens.

APPENDIX C

Classification of words with high front vowels before /l/ from Single Speaker target Word discriminant analyses. Data for C4, U8, and U9 are repeated from Table 6.

		heal	heel	he'll	peal	peel	hill	pill	other
C2	heal	1	3	3	1				
	heel	2	2	1	2				1
	he'll	1	2	2	1	2			
	peal	1	1	1	4	1			
	peel†	2	1	3	1				
	hill						8		
	pill							8	
C4	heal	3	2		2	1			
	heel	2	5			1			
	he'll		1	1			3	3	
	peal	1			6	1			
	peel	4	3			1			
	hill			1			5	2	
	pill						2	6	
U5	heal	4	2	1		1			
	heel†	1		1		4			1
	he'll	1	3	2	1				1
	peal	4	3	1					
	peel	3	1	1	1	2			
	hill						7	1	
	pill							8	
U6	heal		2	1	1	2		1	1
	heel	1		3		3	1		
	he'll	2	2	2	2				
	peal				5	3			
	peel	1		1	1	4	1		
	hill		1		2		4	1	
	pill	1			1			6	
U8	heal		2	3	1	2			
	heel		4	2	1		1		
	he'll		3	3		1	1		
	peal†	1	1		1	1	2	2	
	peel				2	5		1	
	hill	1	2		4	1			
	pill				2		2	4	
U9	heal	2	1	1	2	1			1
	heel	1		3	1	2		1	
	he'll	1	3	3	1				
	peal			3	2	3			
	peel	1	1		1	5			
	hill						4	3	1
	pill						3	5	
U12	heal		3	2	1	2			
	heel†	2		1	1	3			
	he'll	3	1	1	1	1	1		
	peal	3	1	1	1	1	1		
	peel	3	1	2	1	1			
	hill	1					4	1	2
	pill						1	6	1

† indicates fewer than 8 tokens.

APPENDIX D

Classification of words with back vowels before /l/ from Single Speaker VC Rhyme discriminant analyses. Figures represent the proportion of the total number of tokens containing a given vowel that received a particular classification.

		N	/ul/	/ʊl/	/ʌl/	/ɑl/	/ɔl/	/ol/	/ɑp/	/ɑd/	/ɔd/	/op/	/od/	other
C2	/ul/	16	1.00											
	/ʊl/	16		.94				.06						
	/ʌl/	16			.82						.06	.06		.06
	/ɑl/	24				.92				.08				
	/ɔl/	32					.88				.12			
	/ol/	32						1.00						
C4	/ul/	16	.88					.12						
	/ʊl/	16		1.00										
	/ʌl/	16		.06	.82	.06						.06		
	/ɑl/	20			.04	.71		.04	.09					
	/ɔl/	32			.04		.96							
	/ol/	32		.09				.91						
U5	/ul/	16	1.00											
	/ʊl/	16		.69	.19			.12						
	/ʌl/	16			.81	.19								
	/ɑl/	23	.04		.17	.26	.30	.13				.09		
	/ɔl/	32			.03	.03	.91							.03
	/ol/	32			.06	.03		.84				.06		
U6	/ul/	16	.94	.06										
	/ʊl/	16		.44	.13			.44						
	/ʌl/	15		.19	.06		.06	.56					.06	.06
	/ɑl/	23			.08	.38	.29	.08		.08				.08
	/ɔl/	32		.06		.06	.72			.09				.06
	/ol/	32	.03	.35	.16	.03		.45						
U8	/ul/	16	.62	.37										
	/ʊl/	16	.31	.69										
	/ʌl/	16			.75	.19	.06							
	/ɑl/	24	.04		.13	.33	.29	.13				.08		
	/ɔl/	32		.03		.03	.94							.03
	/ol/	32		.03	.06	.03		.81				.03		
U9	/ul/	16	.87					.13						
	/ʊl/	16	.25	.38	.25			.13						
	/ʌl/	16	.06		.81			.13						
	/ɑl/	24				.50	.50							
	/ɔl/	27				.30	.66		.04					
	/ol/	32	.16	.13	.09	.03		.56				.03		
U12	/ul/	16	1.00											
	/ʊl/	16		.31	.38			.25					.06	
	/ʌl/	16		.31	.31			.38						
	/ɑl/	23				.48	.48				.04			
	/ɔl/	32				.41	.56							.03
	/ol/	30	.07	.23	.20	.03		.47						

APPENDIX E

Classification of words with front vowels before /l/ from Single Speaker VC Rhyme discriminant analyses. Information on U8 is repeated from Table 7.

		N	/il/	/ɪl/	/el/	/ɛl/	/æl/	/æp/	/ɛd-p/	/ud/	/ɔl/	other
C2	/il/	40	.98	.02								
	/ɪl/	16		1.00								
	/el/	32			1.00							
	/ɛl/	16				1.00						
	/æl/	16					.94	.06				
C4	/il/	40	.85	.15								
	/ɪl/	16		1.00								
	/el/	32	.03		.97							
	/ɛl/	16		.12		.88						
	/æl/	16				.06	.94					
U5	/il/	40	.95									.05
	/ɪl/	16		1.00								
	/el/	32			.97							.03
	/ɛl/	16				.88			.12			
	/æl/	16				.19	.63	.12				.06
U6	/il/	40	.80	.15	.03					.03		
	/ɪl/	16	.19	.81								
	/el/	32		.03	.91					.06		
	/ɛl/	16			.06	.81	.13					
	/æl/	16				.06	.69	.06			.13	.06
U8	/il/	40	.60	.40								
	/ɪl/	16	.38	.62								
	/el/	31			.65	.35						
	/ɛl/	16		.06	.25	.69						
	/æl/	16					.81	.06			.13	
U9	/il/	40	.98	.02								
	/ɪl/	16		1.00								
	/el/	31		.07	.93							
	/ɛl/	16		.06		.94						
	/æl/	15					.93	.07				
U12	/il/	40	.95	.05								
	/ɪl/	16	.06	.81		.13						
	/el/	32	.03		.97							
	/ɛl/	14				1.00						
	/æl/	16					1.00					

APPENDIX F

Pooled classification of words with back vowels before /l/ from Two Speaker VC Rhyme discriminant analyses. The top part of each panel represents the average of separate classifications based on the other two Utah female speakers, and the bottom part classification based on the Connecticut female speaker.

		/ul/	/ʊl/	/ʌl/	/ɑl/	/ɔl/	/ol/	/op/	/od/	/ud/	/ʊd/	/up/	/ʌd/	/ɑd/	/æl/	other
U6	/ul/	.09	.03					.09	.29			.50				
<U8	/ʊl/	.07	.50	.09			.32	.03								
+	/ʌl/	.18	.03	.03			.65	.03	.03	.03		.03				.03
U9	/ɑl/	.06		.02	.23	.19	.10	.04					.02	.06	.16	.09
	/ɔl/	.03		.08	.13	.30	.02							.11	.17	.14
	/ol/	.19	.03				.75	.01								
U6	/ul/	.68	.13					.16	.06			.06				
<C2	/ʊl/	.06	.19				.69	.06								
	/ʌl/	.19	.06				.50	.06	.06		.13					
	/ɑl/			.24	.24		.14	.10	.05		.05		.10		.05	.03
	/ɔl/			.13	.50	.04	.06					.04		.16		.09
	/ol/	.03	.06				.75	.03			.13					
U8	/ul/	.56		.22			.09		.13							
<U6	/ʊl/	.50	.03	.41			.03		.03							
+	/ʌl/		.06	.54	.22	.10		.06	.03							
U9	/ɑl/	.05	.02	.20	.51	.11	.02	.07								.02
	/ɔl/			.10	.55	.28			.02					.02		.03
	/ol/	.11	.45	.38			.06									
U8	/ul/		.06						.13	.06	.75					
<C2	/ʊl/		.50				.06				.44					
	/ʌl/		.06	.63			.06	.19					.06			
	/ɑl/			.48	.23	.04	.09				.17					
	/ɔl/			.63	.09	.25										.03
	/ol/		.53				.22	.03	.03		.19					
U9	/ul/	.13	.40	.35			.12									
<U6	/ʊl/	.06	.13	.13	.03		.46	.10	.09							
+	/ʌl/	.06		.28	.22		.06	.25	.13							
U8	/ɑl/			.09	.48	.28			.02							.13
	/ɔl/			.03	.42	.43								.02		.10
	/ol/	.06	.03	.14	.02	.02	.26	.31	.16							
U9	/ul/	.06	.38				.31			.06	.19					
<C2	/ʊl/		.56				.13	.19	.13							
	/ʌl/		.31	.13			.13	.44								
	/ɑl/			.63	.25			.04						.08		
	/ɔl/			.67	.19			.13						.13		
	/ol/	.03	.13	.03			.59	.09	.13							

APPENDIX G

Pooled classification of words with front vowels before /l/ from Two Speaker VC Rhyme discriminant analyses. The top part of each panel represents the average of separate classifications based on the other two Utah female speakers, and the bottom part classification based on the Connecticut female speaker.

		/il/	/ɪl/	/el/	/ɛl/	/æl/	/æp/	other
U6	/il/	.25	.38	.27	.09			.02
< U8	/ɪl/	.65	.23	.03	.03			.06
+	/el/	.06		.86	.05			.03
U9	/ɛl/			.06	.72	.22		
	/æl/				.10	.62	.17	.11
U6	/il/		1.00					
< C2	/ɪl/		1.00					
	/el/	.13	.63	.25				
	/ɛl/		.44		.38	.19		
	/æl/				.06	.75		.19
U8	/il/	.55	.39	.06				
< U6	/ɪl/	.45	.55					
+	/el/	.32	.11	.34	.23			
U9	/ɛl/	.22	.06	.13	.60			
	/æl/				.07	.69	.10	.14
U8	/il/	.33	.67					
< C2	/ɪl/	.12	.88					
	/el/		.74	.03	.23			
	/ɛl/		.31		.69			
	/æl/					.81		.19
U9	/il/	.58	.34	.05				.03
< U6	/ɪl/	.47	.06	.28	.19			
+	/el/	.21		.69				.10
U8	/ɛl/	.03		.03	.75	.06		.12
	/æl/				.27	.63		.10
U9	/il/	.13	.80	.08				
< C2	/ɪl/		1.00					
	/el/	.03	.67	.30				
	/ɛl/		.19		.69			.12
	/æl/				.07	.93		

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 113-131

The Role of Fundamental Frequency in Signaling Linguistic Stress and Affect: Evidence for a Dissociation*

Gerald W. McRoberts,[†, ‡] Michael Studdert-Kennedy, and Donald P. Shankweiler[‡]

The fundamental-frequency (*F*0) of the voice is used to convey information about both linguistic and affective distinctions. However, no research has directly investigated how these two types of distinctions are simultaneously encoded in speech production. This study provides evidence that *F*0 prominences intended to convey linguistic or affective distinctions can be differentiated by their influence on the amount of final syllable *F*0 rise used to signal a question. Specifically, a trading relation obtains when the *F*0 prominence is used to convey emphatic stress. That is, the amount of final syllable *F*0 rise decreases as the *F*0 prominence increases. When the *F*0 prominence is used to convey affect, no trading relation is observed.

INTRODUCTION

Speech communications often convey both linguistic content and the speaker's affective state. It is sometimes supposed that linguistic information and affect are independently encoded in the acoustic speech signal, since the same sentence may be spoken with varying affective tones of voice and, in consequence, take on different meanings (e.g., Scherer, Ladd, & Silverman, 1984). Indeed, studies showing that a speaker's intended affect can be identified in languages unfamiliar to the listener, or when the verbal content is removed by filtering, support this assumption (e.g., Kramer, 1964; McCluskey, Albas, Niemi, Cuevas, & Ferrer, 1975; Davitz, 1964; Starkweather, 1961). Further support comes from studies of hemispheric specialization, which suggest different degrees of involvement of the left and right hemispheres for linguistic and affective aspects of speech (e.g., Zurif, 1974; Ley & Bryden, 1982; Shipley-Brown, Dingwall, Berlin, Yeni-Komshian, & Gordon-Salant, 1988; Tucker, Watson, & Heilman, 1977; Weintraub, Mesulam, & Kramer, 1981; Heilman, Bowers, Speedie, & Coslett, 1984; Blumstein & Cooper, 1972). However, little research has directly addressed the issue of how linguistic and affective aspects of speech are encoded during speech production. The purpose of this study was to investigate the relation between linguistic and affective uses of voice fundamental frequency (*F*0).

Speakers use fundamental frequency to convey several linguistic distinctions, including both segmental features such as consonant voicing and vowel height, and prosodic, or suprasegmental features, which typically extend over more than a single segment. Intonation belongs to the prosodic aspect of language; the term refers to variation of *F*0 for linguistic purposes. The acoustic manifestation of intonation is the fundamental frequency of the voice, which contributes to such linguistic distinctions as sentence type. For example, questions and statements are characterized by different fundamental frequency contours in English and in many other languages: Declaratives typically have a gradual decline in *F*0 from beginning to end, while questions (especially syntactically unmarked yes-no questions) tend to have an elevated or rising intonation contour (e.g., Bolinger, 1978; Ultan, 1978), either over the entire utterance, or over its final syllable(s). A further linguistic use of *F*0 in production is contrastive stress, in which one or more words in a sentence may carry added stress

This work is based on the doctoral dissertation of the first author, and was supported by NICHD grant HD-01994 and NIDCD grant DC-00403.

to denote contrastive emphasis. Increases in degree of emphasis are associated with increases in *F*0 (Fry, 1955, 1958; Bolinger, 1958).

Fundamental frequency also conveys paralinguistic information, such as the affective state of the speaker. Few studies of spontaneously-produced affective speech have been done. However, in those cases where spontaneous emotional utterances have been recorded and analyzed, both average *F*0 and *F*0 range are typically increased in comparison with less affectively marked speech (e.g., Williams & Stevens, 1969, 1972, 1981). Affective expressions simulated by actors are consistent with the results obtained for spontaneous speech in showing higher *F*0 for happiness, anger, and sometimes sadness (e.g., Fairbanks & Pronovost, 1939; Fairbanks & Hoaglin, 1941; Williams & Stevens, 1972; see Scherer, 1986 for a review).

Despite the fact that *F*0 is used to convey both linguistic and affective information, there appear to have been no controlled, experimental studies to investigate how linguistic uses of *F*0 may be influenced by the simultaneous use of *F*0 to convey affect. The present study addresses this question by investigating how linguistic and affective influences on *F*0 are encoded in yes-no questions. A dissociation is demonstrated between the use of *F*0 variation in the production of stress contrasts and its use to mark positive and negative affect.

Question Intonation in English

The intuition of traditional phonetics that the distinction between declaratives and questions in English is signalled by a difference in terminal *F*0 glide (e.g., Pike, 1945; Uldall, 1960), specifically a final syllable *F*0 rise for questions and *F*0 fall for declaratives, is only partially supported by data from production. Whereas declarative utterances reliably show a decrease in *F*0 over the duration of the utterance (Pierrehumbert, 1979; Cohen, Collier, t'Hart, 1982; Cooper & Sorenson, 1981), the final syllable *F*0 rise for questions is not obligatory (Cohen, 1972; Fries, 1964). Attempts to relate variation in question contour to syntactic categories (e.g., "wh" questions, tag questions, yes-no questions) have not proved successful (Cohen, 1972). For example, in investigations of yes-no questions produced during radio and television shows (e.g., Fries, 1964; Lee, 1980), only 40% to 55% of the questions had a rising intonation pattern.

Various researchers have suggested that differences in the amount of final rise for questions may be associated with paralinguistic factors, such as the attitude or emotion of the speaker (Crystal, 1969; Lee, 1956, 1960; Jassem, 1972). Thus, Lee (1956, 1960) finds that questions with falling endings tend to have a firm and insistent quality. Crystal (1969) suggests that rising endings are more friendly and interested, and Jassem (1972) finds fall-rise endings to be more friendly, familiar, informal or intimate than rising endings.

These observations suggest a relation between the perceived shape of question contour and judgments of the speakers' attitude or emotion.[1] Unfortunately, these suggestions are not based on controlled experiments. Rather, in most cases judgment of the speaker's attitude was based solely on the investigator's impressions, rather than on judgments by a group of listeners. In addition, since listener judgments often differ from measured *F*0 contours (Hadding-Koch & Studdert-Kennedy, 1964; Jassem, 1972), the relation of the perceived intonation contours to the actual *F*0 contour for a specific attitude or emotion is unclear.

While careful investigation of the use of final rise in various affective contexts might shed light on this issue, it appears that no systematic acoustic and perceptual study of question intonation has been carried out in which speakers used *F*0 to convey a linguistic distinction and simultaneously to convey contrasting attitudes or emotions. The present investigation is intended to clarify the effect of positive and negative affect on question intonation in English.

Trading Relations in Intonation for Yes-No Questions

A series of perceptual experiments by Hadding-Koch & Studdert-Kennedy (1964, 1965 a,b; Studdert-Kennedy & Hadding, 1973) broke new ground by investigating the whole intonation contour of an utterance, rather than just its terminal glide, and by calling both for linguistic judgments of the whole utterance and for psychophysical judgments of the glide. Thus, in their second experiment, Studdert-Kennedy and Hadding (1973) used a speech vocoder to impose parametric variations in *F*0 on the naturally spoken utterance, *November* [no ˈvɛm bɚ]. They constructed a set of 72 contours that included patterns typical of a male speaker's intonation in both Swedish and American English yes-no questions. The principal variations were in the height of the peak *F*0 on the stressed second syllable, and in the extent of the rise or fall in *F*0 on the final syllable. They also prepared sets of pulse-train analogs and sine-wave analogs with *F*0 contours identical to those of the speech stimuli.

These stimuli were presented in random test orders to groups of Swedish and American listeners, who made both linguistic judgments of the whole utterance (question/statement) and psychophysical judgments of the terminal glide (rise/fall) on the full speech stimuli, and for psychophysical judgments of the terminal glide on the sine-wave and pulse-train analogs. The results on the linguistic judgments confirmed the findings of their previous study (with a different utterance) in two respects. First, for both groups of listeners, terminal glide was the single most powerful determinant of question/statement judgments: all contours judged as *question* over 90% of the time had a rising terminal glide. Second, for both groups of listeners, the height of the peak *F*0 on the medial stressed syllable also affected linguistic judgments: a peak of 200 Hz, starting from an initial *F*0 of 130 Hz, required less terminal rise than a peak of 160 Hz for listeners to judge the utterance a *question*. Studdert-Kennedy and Hadding dubbed this effect a reciprocal *trading relation* between stress peak and terminal rise.

The psychophysical judgments of the full speech stimuli were less consistent than the question/statement judgments, for both groups of listeners. Overall, contours heard as terminally rising were judge to be questions; but not all contours judged to be questions were heard as terminally rising. This asymmetry arose because the effect of the peak *F*0 was less consistent both within and between subjects for the psychophysical than for the linguistic judgments. Moreover, the effect of the peak *F*0 was entirely absent from judgments of the terminal glide on the sine-wave and pulse-train analogs. The authors therefore concluded that the peak *F*0-final rise trading relation was specifically linguistic in its origin.

If this is so, we might suspect that listeners' perceptual systems are attuned to properties that normally obtain in speech, and a parallel trading relation might be found in speech production; i.e., as peak *F*0 is raised to convey stress or emphasis in a question, the amount of final rise should be reduced. In fact, an account of the trading relation based on physiological constraints on speech production was proposed by Lieberman in 1967, but the matter has not been pursued further.

The importance of specifying the perception-production link should not be underestimated. The perceptual studies of Studdert-Kennedy and Hadding (1964, 1973) used hybrid stimuli formed by imposing various synthetic intonation contours on a single utterance. In making these stimuli, the investigators manipulated *F*0 without regard to the constraints of natural speech production. For example, although the range of *F*0 values used in constructing the experimental intonation contours was based on spectrographic data from natural productions (Hadding-Koch, 1961), the combinations of *F*0 values were arbitrary. Moreover, all the stimuli had the same overall duration. As a result, the stimulus set was artificially constrained and much of the variability in frequency spectrum, amplitude envelope, duration and *F*0 characteristic of naturally produced speech was absent in these hybrid stimuli.

The first goal of the present study was then to determine whether the relation noted in the Studdert-Kennedy and Hadding perceptual studies normally obtains in naturally spoken utterances by testing for a peak *F*0-final rise trading relation in speech production. To pursue this goal, Experiment I determined whether a trading relation occurred when the peak *F*0 on the syllable preceding a question final rise was elevated to convey contrastive stress. The expected effect was observed, and Experiment II therefore explored whether the trading relation occurred also when *F*0 was raised for a reason other than stress placement; i.e. in conjunction with variation in affective tone. We predicted that it would not and our prediction was confirmed.

A more stringent test of the relation between peak *F*0 and final rise among contrastively stressed and affectively-toned productions would examine only that subset of tokens for which listener judgments coincided with the speakers' intent. Therefore, Experiment III elicited both question-statement and emotional polarity judgments for the utterances produced in Experiment II, and then reanalyzed the data of that experiment for just those tokens that were perceived by naive listeners in agreement with speaker intent.

EXPERIMENT I

The purpose of this experiment was to explore the influence of peak *F*0 on the amount of final rise in the production of questions. Specifically, this study attempted to confirm by analysis of male speakers' productions the conclusions of Studdert-Kennedy and Hadding that a trading relation exists between peak *F*0 and final rise, such that *F*0 peaks at 200 Hz require less final rise than tokens with peak *F*0 of 160 Hz or below. Linguistically relevant variations in peak *F*0 were induced by having speakers produce questions with varying degrees of contrastive stress and without contrastive stress. We hypothesized that

variations in peak *F*0 resulting from differences in contrastive stress would influence the amount of final rise used by speakers when producing questions—in other words, we predicted a trading relation between the peak *F*0 and the amount of final rise. To test this hypothesis, acoustic analysis was performed to determine the relation between second-syllable peak *F*0 and final-syllable *F*0 rise.

Methods

Subjects. Four adult males, native speakers of American English and ranging in age from 23 to 35 years were the talkers. All were from the northeastern or midwestern regions of the United States and without marked regional accents.

Test Utterances. Two Utterances were used in this study: *November* and *I did it.* November was included to enable a direct comparison with the findings of Studdert-Kennedy and Hadding (1973). The sentence *I did it* was included for the following reasons: 1) it is meaningful as both question and statement, both with and without contrastive stress; 2) it is voiced throughout; 3) it serves as a phonetic control on *November* by eliminating the difference in vowel quality between the second and third syllables and the nasalization and frication, which might interact with *F*0; 4) its three constituent syllables are independent morphemes, as opposed to the three syllables in *November* which are not. Only the questions are considered in the following analyses.

Procedure. The experimenter and each talker were seated in a sound booth, with the talker positioned approximately 24 inches from a microphone. The microphone provided input to a remotely controlled tape recorder in an adjacent sound booth. Brief descriptive scenarios provided contextual support for the subject in each condition. For example, in one scenario the speaker responds to a young child who says that Christmas is in November. The response is a contrastively stressed question—"November?... Don't you mean December?" The *November* scenarios were described by the experimenter and each talker produced his responses to each scenario. The experimenter, without resorting to models or examples, encouraged the talkers to vary emphasis over a range of values in the contrastive stress condition. The procedure was repeated for the *I did it* scenarios.

Each speaker produced the two sentences in each of three conditions: 1) declarative (statement) with neutral stress; 2) question with neutral stress; 3) question with contrastive stress. Each speaker produced 4-5 repetitions of each sentence in each condition for a total of approximately 30 tokens per subject and 120 total tokens, of which one-third were statements and the two-thirds subjected to analysis were questions.

Acoustic Analysis. Each talker's recorded tokens were digitized at a 10 KHz sampling rate using the Haskins Laboratories PCM (Pulse Code Modulation) system. Input levels were held constant across each talker's entire session. The digitized tokens were stored on disk. Individual tokens were prepared for analysis using the Wave Form Editing and Display (WENDY) software at Haskins Laboratories. Fundamental-frequency analysis was performed using Interactive Laboratories Systems (ILS) software (Signal Technology, Inc., 1978). The parameter settings for initial analysis of all tokens included a 10 millisecond sampling window with a 50% overlap between adjacent windows, minimum and maximum *F*0 values of 75 Hz and 400 Hz, respectively, and a voicing threshold value of -400. (The sampling window and overlap parameters provide for some smoothing of the cycle-to-cycle variation in *F*0, especially at higher frequencies. The voicing threshold setting increases the likelihood that periodicity will be found throughout the token.)

The result of the acoustic analysis for each token was displayed in the form of an *F*0 contour, where each value in the contour represented an analysis frame corresponding to a 10 millisecond sampling window (Figure 1). The contours were inspected to insure that the analysis provided plausible *F*0 values. When suspect values were encountered, the token was re-analyzed using *F*0 values obtained by counting pitch pulses in the waveform display as a guide to resetting the ILS analysis parameters.

The following measurements were taken from the contour of each token: 1) second-syllable peak *F*0; 2) *F*0 at third-syllable onset; 3) third-syllable peak *F*0. The amount of final rise was defined as the difference between *F*0 at the third-syllable onset and the third-syllable peak *F*0. Syllable onset was determined by inspecting the *F*0 contour for a down-turn which resulted from closure (for /b/ or /d/) at the boundary between the second and third syllables (see Figure 1). The frame containing the lowest point in the down-turn was taken as the onset of the third syllable. On the few occasions in which the *F*0 down-turn did not occur, the waveform was inspected and the average of the pitch periods over the first 10 milliseconds of the third syllable, beginning with the first pitch period following the release of the /b/ or /d/, was used as the value of *F*0 at onset.

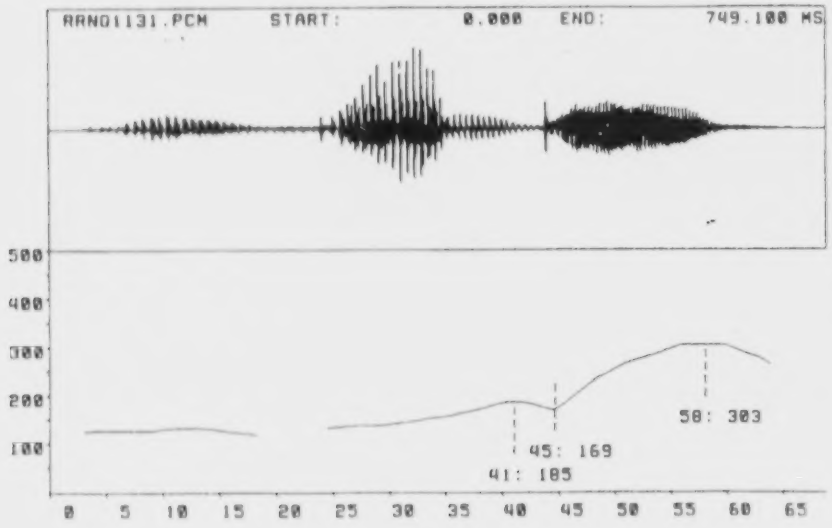


Figure 1. **Fundamental-frequency contour for *November* indicating values at second-syllable peak F0 (frame 41: 185 Hz), third syllable onset (frame 45: 169 Hz) and final syllable peak (frame 58: 303 Hz).**

Results

Median peak F0 and final rise. The distributions of peak *F0* and final rise values for each speaker in the Sentence × Stress Conditions were examined and typically found to be skewed or polymodal. Therefore, median peak *F0* and final rise values were determined for each of the four speakers in each condition. These values are presented in Table 1. Mean median peak *F0* is higher for contrastive than for neutral stress in both utterances, but there seems to be no systematic effect of stress on the final rise.

Separate two-way ANOVA's (Sentence × Stress) were performed on these median peak and final rise values. As expected, this analysis indicated that median peak *F0* was significantly higher for questions with contrastive stress than for questions with neutral stress (210.7 Hz vs. 140.5 Hz; $F(1,3)=102.79$, $p<.001$). This was consistent for all four speakers on both utterances. Median peak *F0* did not differ between the utterances (172.4 vs. 186.2; $F(1,3)=0.09$) and the Stress × Utterance interaction was not significant ($F(1,3)=0.79$). The analysis of final rise indicated no significant effects and no significant interactions.

Peak F0-Final Rise Relation. Correlation coefficients (Pearson r) between peak *F0* and final rise were computed for the 4-5 tokens produced by each speaker in each condition and are shown in Table 2. In the Neutral Stress Condition, mean correlations ranged from r=-.52 to r=.80, while in the Contrastive Stress Condition all correlations were negative, ranging from r=-.60 to r= -.75.

Table 1. *Median Peak F0 and Final Rise in Hz.*

	November				*I did it*			
	Neutral		Contrastive		Neutral		Contrastive	
Speaker	Peak	Rise	Peak	Rise	Peak	Rise	Peak	Rise
L R	126.5	108.5	172.0	98.0	148.0	90.5	204.0	135.0
R M	159.5	78.5	278.5	73.5	129.5	72.0	159.0	188.0
J S	159.0	74.0	211.0	67.5	162.5	116.5	256.0	81.0
G B	123.0	93.5	204.2	111.0	116.0	81.0	202.0	143.5
Mean Median	142.0	88.6	216.4	87.5	139.0	90.0	205.2	136.9

Table 2. *Correlations between Peak F0 and Final Rise for Two Utterances Produced with Neutral and Contrastive Stress.*

	November		*I did it*		*Mean*	
Speaker	Neutral	Contrastive	Neutral	Contrastive	Neutral	Contrastive
L R	.53	-.54	-.86	-.65	-.34	-.60
R M	.88	-.28	.69	-.93	.80	-.75
J S	-.36	-.88	.70	-.44	.39	-.73
G B	-.53	-.84	-.52	-.22	-.52	-.62
M	.41	-.70*	-.04	-.65*	.19	-.68†
SE	(.40)	(.26)	(.54)	(.31)	(.40)	(.07)

*p <.10, one-tailed. †p <.05, one-tailed.

Mean correlation coefficients across utterances for each speaker, and across speakers, were obtained by converting each speaker's correlation coefficient in each condition to z_r using Fisher's transform (Ferguson, 1981). The z_r values were then averaged and re-converted to r's. Since each speaker produced only 4-5 tokens per Sentence × Stress Condition, individual speaker's mean correlations were not tested for significance. The significance of the deviation of the group mean r's from zero was tested for each utterance and for the average of the two utterances by t-tests for samples of N=4, df=3. Tests of significance in the Contrastive Stress Condition were one-tailed, since the hypothesis was that these correlations would be negative. In the Neutral Stress Condition, tests of significance were two-tailed.

Group mean correlation coefficients (again determined through Fisher's transformation) are shown with the standard errors of the means in parentheses in the bottom row of Table 2. In the Neutral Stress Condition, the mean correlation is near zero for *I did it* and positive for *November*. However, in the Contrastive Stress Condition significant negative correlations occurred for both utterances. The overall mean correlation (across utterances and speakers) in the Contrastive Stress Condition was r=-.68 (p <.001) and in the Neutral Stress Condition the overall mean correlation was r=.19 (ns). The peak and final rise values for each speaker's tokens in each Sentence × Stress Condition were converted to z-scores and are plotted in Figure 2 for *November* and *I did it* questions, respectively.

Intonation contour shape. The shape of the intonation contours used by the speakers was generally the continuously rising form typical of American-English Yes-No questions (e.g., Pike, 1945). All tokens of each speakers' Neutral Stress questions were of this form, as were most of their tokens of Contrastive Stress questions (See Figure 3). However, two speakers (JS & RM) used rise-fall-rise contours for some of their Contrastive Stress questions. Speaker JS used this pattern for two Contrastive Stress tokens of "November."

Discussion

The results of this experiment support the hypothesis that the trading relation observed by Studdert-Kennedy and Hadding (1973; Hadding-Koch & Studdert-Kennedy, 1965a,b) in the perception of question intonation also occurs in production. A significant negative correlation was found between stressed syllable peak *F0* and the amount of final rise for questions produced with contrastive stress. The overall group mean correlation for questions with contrastive stress was r=-.68. The group mean correlations were of similar magnitude for each of the two utterances *November* and *I did it* (r=-.70 and r=-.65, respectively), suggesting that the segmental make-up of the utterance is probably not a major determining factor. The occurrence of the trading relation in the contrastive stress condition was consistent across speakers: negative correlations were found for all four speakers.

For questions with neutral stress, the mean peak *F0*-final rise correlation did not differ significantly from zero (r=.19) overall or for either utterance individually. For these questions, individual speakers varied from a strong negative to a strong positive peak-final rise relation. This variability may be due to the limited range of peak *F0* values for questions without contrastive stress, or it may reflect the fact that the trading relation is related specifically to contrastive stress.

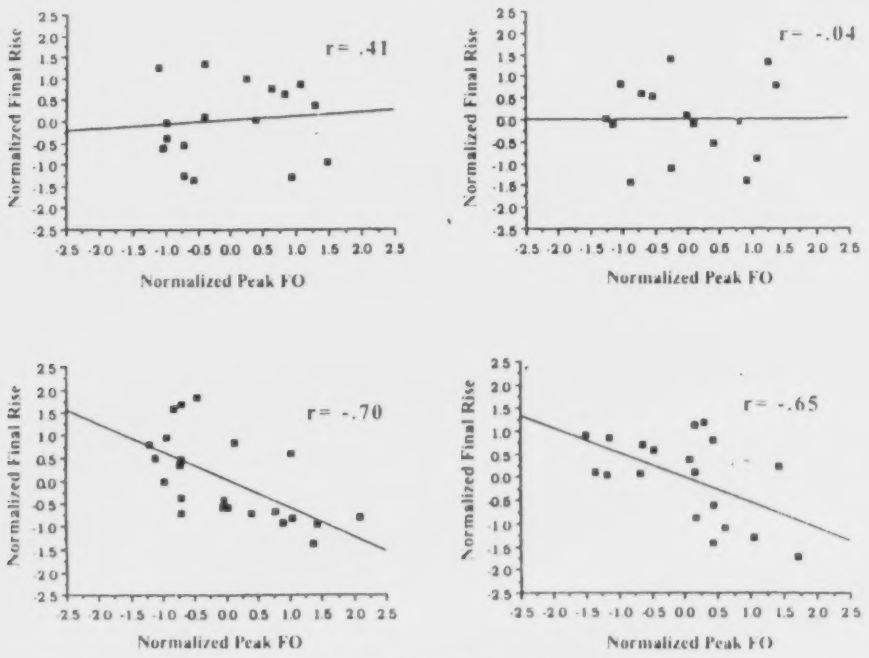


Figure 2. **Relations between peak *F0* and amount of final rise for *November* and *I did it* questions produced with neutral (top) and contrastive stress (bottom).**

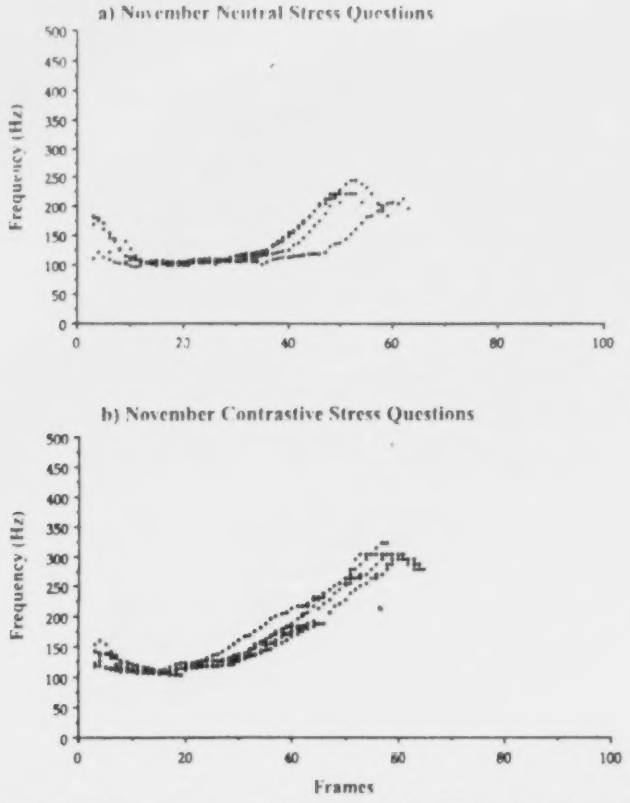


Figure 3. **Continuously rising intonation contours produced by speaker LR for the utterance "November."**

BEST COPY AVAILABLE

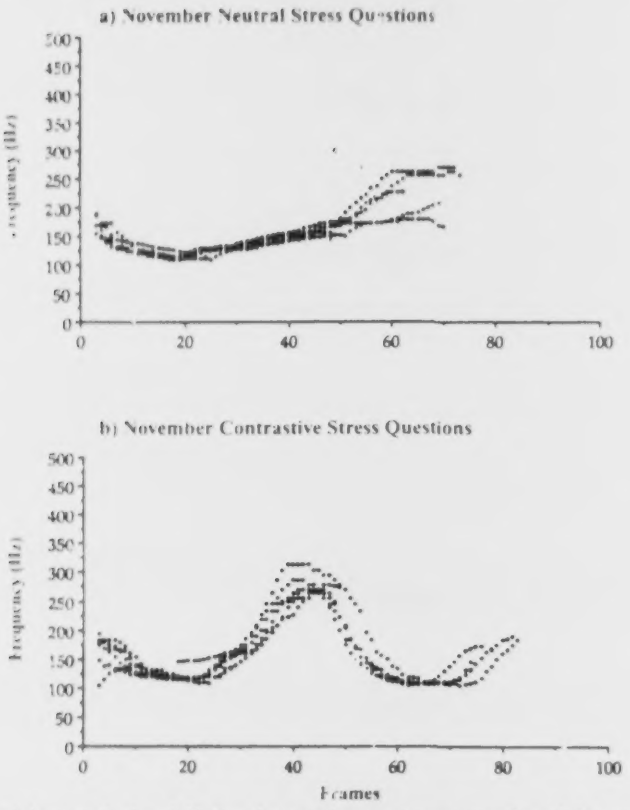


Figure 4. Continuously rising and rise-fall-rise intonation contours produced by speaker JS or the utterance "November."

A potential concern with the results in the Contrastive Stress conditions is that although most of the tokens in this condition had continuously rising contours, two speakers produced some tokens with rise-fall-rise contours. If the peak *F*0-final rise relation differs for different contour shapes, combining data across these patterns could lead to invalid conclusions. Thus, for example, if the amount of final rise for rise-fall-rise contours was generally less than for continuously rising contours, then the finding of a negative peak *F*0-final rise correlation could be spurious. This seems unlikely for two reasons. First, note that all the correlations in the Contrastive Stress condition for every subject were negative (see Table 2). Thus, it is not the case that subjects who produced variant contour shapes were spuriously causing the negative correlation in the Contrastive Stress condition. Second, the correlations for subjects that used the rise-fall-rise pattern were not consistently in a direction to influence the overall pattern of results. For example, speaker JS, who used the rise-fall-rise contour for all of his Contrastive Stress tokens, had peak-final rise correlations of r=-.88 and r=-.44. And speaker RM, who produced two tokens with rise-fall-rise contours for the "November" utterance had a correlation of r=-.28 for that utterance, but had a correlation of r=-.94 in the "I did it" condition that included only continuously rising contours. Thus, our finding of a negative correlation between peak *F*0 and final rise for Yes-No questions with contrastive stress does not appear to be influenced by the presence of multiple contour shapes in the data.

The detailed pattern of results in this experiment corresponds closely to that obtained by Studdert-Kennedy and Hadding (1973) for the perception of question intonation. For example, changes in peak *F*0 between 130 and 160 Hz had no consistent effect in perception, but changes from 160 to 200 Hz reliably reduced the amount of final rise listeners required to judge an utterance to be a question. In the current investigation, the mean median peak *F*0 in the neutral stress condition was approximately 140 Hz for each of the two utterances (*November* and *I did it*) and no trading relation was found. On the other hand, in the con-

trastive stress condition, the mean median peak *F*0 was greater than 200 Hz and a trading relation was found. The close parallel between the present results in production and those of Studdert-Kennedy and Hadding in perception strongly suggests that the perceptual trading relation reflects listeners' attunement to patterns of speech production. This is consistent with the interpretation by Studdert-Kennedy and Hadding (1973) that their listeners perceived increases in peak *F*0 above 160 Hz as due to contrastive or emphatic stress, but does not rule out the possibility that the trading relation occurs whenever peak *F*0 increases to a high value. Experiment II addressed this possibility by investigating whether the trading relation occurs when peak *F*0 increases to convey a difference in affect.

EXPERIMENT II

The goal of this experiment was to replicate the trading relation found in Experiment I and explore its limiting conditions. It will be recalled that in their perceptual studies, Studdert-Kennedy and Hadding found that the influence of peak *F*0 on the amount of final rise was limited to linguistic (question-statement) judgments. It was not found for judgments of the direction of pitch change (terminal rise or fall). Studdert-Kennedy and Hadding interpreted the dissociation between question-statement and rise-fall judgments with regard to the influence of peak *F*0 as evidence that the trading relation was a linguistic phenomenon, not a general psychoacoustic effect. Accordingly, in the present study it was hypothesized that in production a trading relation would occur when peak *F*0 varied to convey a linguistic distinction, but not when it varied to convey an affective state. Trained actors were engaged to produce questions (and statements) with and without contrastive stress, as in Experiment I. In addition, the same actors produced utterances without contrastive stress, but conveying positive and negative affects. It was predicted that the trading relation between peak *F*0 and amount of final rise in question intonation would occur in the contrastive stress condition, but not in the affective conditions.

Methods

Subjects. Trained speakers were used as subjects. The speakers were four male actors, aged 23-28 years, recruited from the Yale Drama School. All were native speakers of American English from the northeastern and mid-Atlantic region. Each speaker was paid for his participation in a single 1 1/2 - 2 hour session.

Test utterances. The test utterances were those used in Experiment I: *November* and *I did it*.

Procedure. A general description of the experiment was given to each speaker, followed by a 3-5 minute calibration period (Cosmides, 1983) during which each speaker read a dramatic script from a science fiction novel (LeGuinn, 1968). Speakers were provided with written descriptions of the scenarios and given 10-15 minutes to study them and to prepare their productions. They were instructed to make four or five repetitions as nearly identical as possible for each scenario and to use facial gestures appropriate for each affect condition, since speaking with and without smiling has been shown to result in acoustic differences which listeners readily perceive (Tartter, 1980). Each speaker was seated in a sound booth approximately 24 inches from a microphone and a video camera. In an adjacent sound booth the audio channel was recorded on a tape recorder and the video channel was recorded on a VHS video recorder.

Each speaker produced the two sentences in a set of eight Stress Conditions and a set of eight Affect Conditions. The Stress Conditions resulted from crossing two degrees of stress (neutral and contrastive) with two sentence types (statement and question) and two listener-speaker distances (near and far). In the last condition, speakers were instructed to speak first at an ordinary conversational level and then to raise the output level, as if to speak up over a short distance (e.g., across a table) or over moderate background noise; this condition was included to ensure that the range of peak *F*0 in the Stress Conditions was not unnaturally constrained. The Affect Conditions result from crossing two affects (positive and negative), with two sentence types (question and statement) and two degrees of intensity (mild and moderate), was again intended to ensure a full range of peak *F*0 values. As in Experiment I, brief descriptive scenarios provided contextual support for each condition. Each speaker produced 4-5 repetitions in each condition for a total of approximately 128 tokens per subject and 512 total tokens, of which half were statements and half were questions. In the following analyses only the 256 questions are considered and the data are collapsed across the near-far and mild-moderate dimensions of the design providing 8-10 trials for each subject in each condition.

Subjects completed all Stress Conditions before going on to the Affect Conditions. Within the Stress Conditions, all the *November* tokens were spoken first (neutral-contrastive statements,

neutral-contrastive questions), followed by the *I did it* tokens in the same order of conditions. Within the Affect Conditions, positive affect tokens of both test sentences were spoken first in the order statements-questions, then negative affect tokens of both test sentences in the same order. After the Stress Conditions, and after the Affect Conditions, speakers reviewed the tapes of their productions. If a speaker expressed dissatisfaction with any of his productions, an opportunity was provided to produce additional tokens.

Acoustic analysis. The data were treated as in Experiment I. Thus, the following measurements were made on each token: 1) second-syllable peak *F*0; 2) *F*0 at third-syllable onset; 3) third-syllable peak *F*0. The amount of final rise was defined as the difference between *F*0 at third-syllable onset and the third-syllable peak *F*0. Third-syllable onset was determined as in Experiment I.

Results

Median peak F0 and final rise. As in Experiment I, median peak *F*0 and final rise values for each speakers' questions were determined and are shown for the Stress Conditions in Table 3. Mean median peak *F*0 was greater for tokens with contrastive stress than for those with neutral stress. This was the case for each speaker and each sentence, with one exception (speaker PN's median peak *F*0 is about equal for *November* tokens with neutral and contrastive stress). Mean median final rise is greater for tokens with contrastive stress than for those with neutral stress for the sentence *November*, but the reverse is true for the sentence *I did it*. Two-way ANOVA's, Utterance (November vs. I did it) × Condition (Neutral Stress vs. Contrastive Stress or Postive vs. Negative), were performed separately on the median peak *F*0 and final rise data for the Stress Conditions and the Affect Conditions. For peak *F*0 in the Stress Conditions, the main effect of Stress was significant ($F(1,3)=14.92$; $p < .05$), reflecting the fact that median peak *F*0 was higher in the Contrastive Stress Condition than in the Neutral Stress Condition. The main effect of Utterance was not significant ($F(1,3)=4.98$; $p > .10$), nor was the Condition × Utterance interaction ($F(1,3)=0.04$). For final rise, a two-way ANOVA, Utterance (November vs. I did it) × Condition (Neutral Stress vs. Contrastive Stress) indicated no significant main effects and no significant interactions.

Median peak *F*0 and final rise values for each speakers' questions in the Affect Conditions are shown in Table 4. Mean median peak *F*0 is greater for positive than negative affect for the sentence *I did it*, but for the sentence *November*, the mean median peak *F*0 is slightly greater for negative affect tokens than for positive affect tokens. Mean median final rise is greater for positive than for negative affect tokens for both utterances, a result that is seen for all but one speaker (speaker RR produced slightly greater final rise for negative than positive affect tokens of *I did it*). A two-way ANOVA (Utterance × Affect), performed on median peak *F*0 indicated no significant effects and no significant interactions. For Final Rise, the ANOVA indicated a main effect for Utterance ($F(1,3)=13.29$; $p < .05$), reflecting more final rise for *I did it* questions than for *November* questions. No other main effects or interactions were significant.

Table 3. *Median Peak F0 and Final Rise for Stress Conditions in Hz.*

	November				*I did it*			
	Neutral		Contrastive		Neutral		Contrastive	
Speaker	Peak	Rise	Peak	Rise	Peak	Rise	Peak	Rise
EO	104.5	60.0	153.0	40.5	131.0	48.0	146.5	45.0
PN	124.5	62.5	123.5	69.0	118.5	112.5	136.0	48.5
RR	133.5	75.7	164.0	149.9	162.0	112.0	173.5	101.5
SB	110.0	95.0	135.5	52.0	101.0	48.0	149.0	96.0
Mean Median	118.1	73.3	144.0	77.8	128.1	80.1	152.2	72.7

Table 4. *Median Peak F0 and Final Rise for Affect Conditions in Hz.*

	November				*I did it*			
	Positive		Negative		Positive		Negative	
Speaker	Peak	Rise	Peak	Rise	Peak	Rise	Peak	Rise
EO	129.5	48.5	137.5	12.5	138.5	45.5	134.0	26.0
PN	260.5	72.0	256.5	65.0	219.5	140.0	133.5	23.5
RR	200.0	77.0	192.0	106.0	263.0	94.0	186.0	97.0
SB	139.0	53.5	171.5	48.0	137.0	98.0	169.5	57.5
Mean Median	182.2	62.7	189.4	57.9	189.5	94.4	155.7	51.0

Peak F0-final rise relation. Correlations between peak *F*0 and final rise for the 8-10 tokens produced by each speaker in each of the four Stress and each of the four Affect Conditions are shown in Tables 5 and 6, respectively. Mean correlation coefficients were computed across conditions within speaker and across speakers using Fisher's transform. The significance of the deviation of the group mean r from zero was tested on samples of N=4, df=3. As in Experiment I, one-tailed significance tests were performed in the Contrastive Stress Condition and two-tailed tests were performed in all other conditions.

For individual speakers, mean correlations across the test sentences in the Contrastive Stress Condition ranged from r=-.42 to r=-.78, while the correlations in the Neutral Stress Condition ranged from r=-.11 to r=.35 (see Table 5). In the Affect Conditions, mean correlations for individual speakers averaged across test sentences ranged from r=-.36 to r=.83 for the two affects separately (see Table 6). Overall mean correlations for individual speakers averaged across the test sentences and affects ranged from r=-.27 to r=.55. In summary, every speaker demonstrated negative mean correlations in the Contrastive Stress conditions, while correlations in the Neutral Stress and Affect Conditions ranged over positive and negative values.

The group mean correlations between peak *F*0 and final rise in the Stress Conditions appear with their standard errors in parentheses in the bottom row of Table 5. In the Neutral Stress Condition the overall mean correlation was r=.15, which did not differ significantly from zero. For the two utterances separately, both correlations were positive, but neither was significant. However, in the Contrastive Stress Condition, the overall mean correlation was r=-.67 (p <.001), and significant negative correlations were found for each test sentence.

Table 6 presents the corresponding correlational data for the Affect Conditions. The mean peak *F*0-final rise correlation across speakers, test sentences, and affects was r=.34, which is not significantly different from zero. Mean correlations across the two Affect Conditions for each of the test sentences were r=.44 (ns) and r=.05 (ns) for *November* and *I did it*, respectively. Mean correlations across test sentences for the Positive and Negative Affect Conditions were r=-.11 (ns) and r=.52 (ns), respectively. For the two test sentences individually, mean correlations were r=.33 and r=.55 for *November* and r=-.57 and r=.50 (all ns) for *I did it* in Positive and Negative Affect Conditions, respectively. Thus, across subjects in the Affect Conditions there were no correlations which differed significantly from zero.

Table 5. *Correlations between Peak F0 and Final Rise for Each Speaker in Each Stress Condition.*

	November		*I did it*		*Mean*	
Speaker	Neutral	Contrastive	Neutral	Contrastive	Neutral	Contrastive
EO	.18	-.65	.05	-.78	.11	-.72
PN	-.20	-.51	-.03	-.33	-.11	-.42
RR	.25	-.84	.27	-.47	.26	-.70
SB	.69	-.94	-.12	-.33	.35	-.78
M	.26	-.79**	.04	-.51**	.15	-.67†
SE	(.22)	(.26)	(.08)	(.17)	(.10)	(.13)

*p <.05, one-tailed. †p <.01, one-tailed.

Table 6. *Correlations between Peak F0 and Final Rise for Each Speaker in the Affect Conditions.*

	November		*I did it*		*Mean*		
Speaker	Positive	Negative	Positive	Negative	Positive	Negative	Overall Mean
E O	-.23	-.34	-.41	-.10	-.32	-.23	-.27
P N	.37	.83	-.25	.83	.06	.83	.55
R R	.72	.88	-.93	.66	-.36	.79	.34
S B	.31	.25	.07	.30	.19	.28	.24
M	.33	.55	-.57	.50	-.11	.52	.34
SE	(.23)	(.41)	(.37)	(.28)	(.14)	(.34)	(.09)

The peak *F0* and final rise values for each speaker in each Sentence × Stress and Sentence × Affect Condition were converted to z-scores and are plotted in Figures 5 and 6 for the Stress and Affect Conditions respectively.

Intonation Contour Shapes. All tokens included from all four speakers in both the Neutral Stress and Contrastive Stress conditions had the continuously rising pattern typical of American-English Yes-No questions. In the affect conditions, all of the Positive Affect contours and most of the Negative Affect contours were also of the continuously rising form. In the Negative Affect condition, some tokens produced by speakers EO and SB had contours that might best be described as rise-flat, or rise-fall-flat, while speaker PN produced tokens with the rise-fall-rise contour for the utterance "November" in this condition.

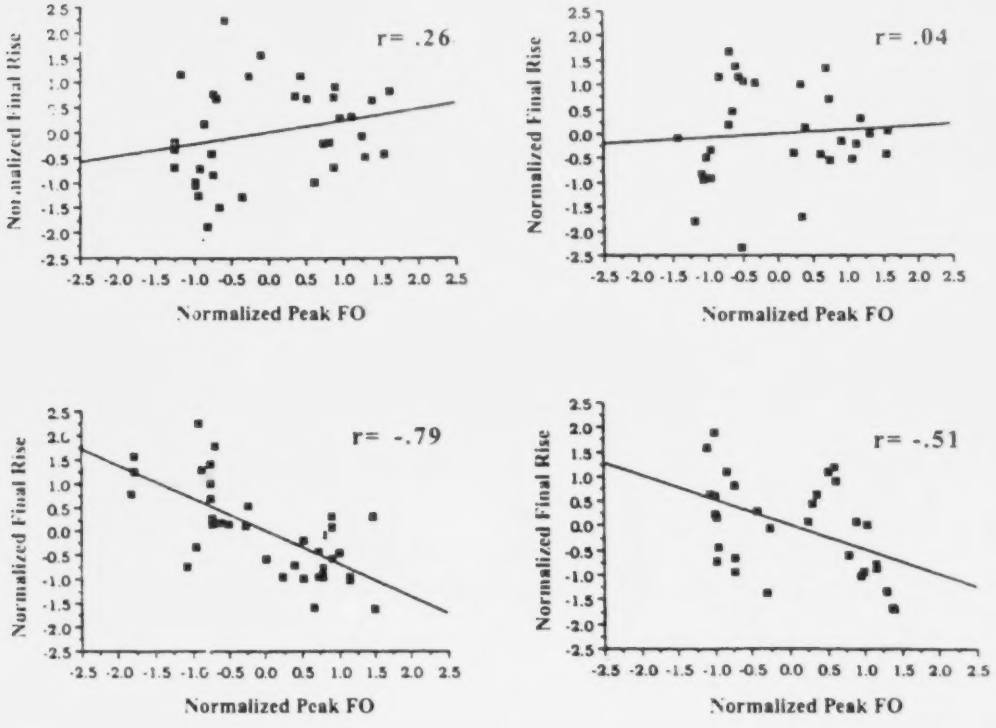


Figure 5. **Relation between peak *F0* and amount of final rise for *November* and *I did it* questions produced with neutral (top) and contrastive stress (bottom).**

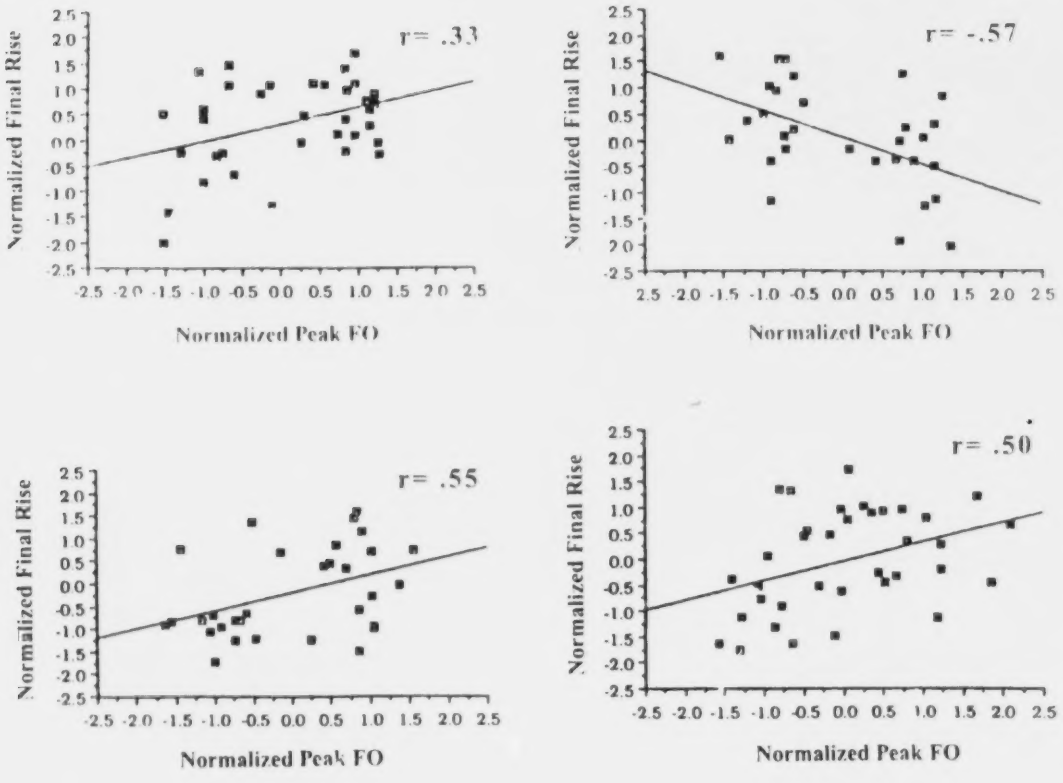


Figure 6. **Relation between peak *F*0 and amount of final rise for *November* and *I did it* questions produced with positive (top) and negative affect (bottom).**

Discussion

The purpose of Experiment II was to confirm the findings of Experiment I and to investigate whether the trading relation would also occur when peak *F*0 varied to convey affect. To that end, the experiment included the same test sentences and the same Neutral and Contrastive Stress Conditions as Experiment I. In addition, conditions were added in which speakers conveyed positive and negative affect.

The results of this experiment replicate the findings of Experiment I. The trading relation between peak *F*0 and final rise occurred among questions with contrastive stress, where significant negative mean correlations were found between peak *F*0 and final rise overall (r=-.67; $p < .001$), and for each test sentence (r=-.79, $p < .05$; r=-.51, $p < .05$). However, among questions without contrastive stress, the overall correlation was nonsignificant (r=.15).

In the Affect Conditions no evidence of a trading relation was found. Rather, the peak *F*0-final rise relation among questions conveying affect was characterized by an overall mean correlation which did not differ significantly from zero (r=.34). Since the trading relation did not occur in the Affect Conditions despite a higher mean peak *F*0 among these affectively-toned questions than among questions with contrastive stress, it may be seen that the effect is not a result of raising peak *F*0 generally. Instead, the trading relation appears to be specific to the use of *F*0 to convey contrastive stress.

As in Experiment I, these results are clear in their indications. Each speaker demonstrated negative correlations in the Contrastive Stress Condition for both utterances (r=-.33 to r=-.94), while in the Neutral Stress Condition, correlations varied from small negative to moderate positive (r=-.20 to r=.69). Similarly, in the Affect Condition, speakers' overall correlations ranged from small negative to moderate positive (r=-.27 to r=.55). Considering the two emotions separately for each speaker, five out of eight mean correlations were positive.

These results demonstrate a dissociation in the trading relation between linguistically driven and affectively driven variations of *F*0. That is, when *F*0 is increased to convey contrastive stress, the amount of final rise used to signal a question decreases. However, when *F*0 increases to convey affect, the amount of final rise tends to increase. This dissociation supports the hypothesis that the trading relation is specifically a linguistic phenomenon. This specificity raises the possibility that the articulatory actions used to raise *F*0 for contrastive stress may differ from the actions that raise *F*0 to convey positive and negative affects. In short, elevation of *F*0 for contrastive stress appears to interfere with the production of the final syllable rise in *F*0, whereas elevation of F0 to convey affect does not impede production of final rise.

EXPERIMENT III

This experiment seeks to validate the results of Experiment II by ascertaining that naive listeners could apprehend the speakers' intent. Therefore, in Experiment III listeners were asked to rate each token from Experiment II as either a question or statement. The affective polarity and degree of affect of each token were also rated. Of special interest was whether the differential occurrence of the trading relation would be maintained in those specific tokens for which listeners correctly apprehended the speaker's intentions. The subset of tokens for which the majority of listeners apprehended the speaker's intent with regard to question-statement and affect were examined for the relation between peak *F*0 and final rise.

Methods

Subjects. Eighteen adult native English speaking subjects participated in Experiment III. Subjects were graduate and undergraduate students from the University of Connecticut. All subjects indicated they had normal hearing.

Stimuli. The stimuli were all of the *November* questions and statements produced by each of the four speakers in Experiment II. Each speaker's productions were randomized separately and arranged in blocks of ten tokens with seven-second interstimulus intervals and 15-second interblock intervals.

Procedure. Subjects were tested individually in a sound attenuated room. Each subject heard the productions of two speakers in separate tests with a five minute break between tests. Before each test, subjects were instructed that they would hear a single speaker saying the word *November* as either a question or statement and in various tones of voice. Each token was rated as a question or statement and then rated for the polarity and degree of affect on an eleven point scale (-5 to +5), where -5 indicated very negative and +5 indicated very positive. Each token was presented once without feedback. Stimuli were presented over headphones at a comfortable loudness level. Each session lasted approximately 40 minutes.

Results

Those tokens correctly labelled by the majority of listeners as questions and having an average rating consistent with the speaker's intended affect were selected for further consideration. The criterion for inclusion of tokens rated on the affective scale was an average rating of +1.5 to +5 for positive affect, -1.5 to -5 for negative affect, and +1.0 to -1.0 for tokens from the non-affective stress conditions. In general, the listeners' ratings were in good agreement with the speakers' intent. Question-statement judgments were in very high agreement with the speakers' intent, with 98% of the questions and 100% of the statements rated appropriately by the majority of listeners. In addition, 50% of the questions from the neutral and contrastive stress conditions, 62% of the positive and 75% of the negative affect questions met the affective rating criteria for inclusion in the following analysis.

Peak F0-final rise relation among questions. The mean correlations between peak *F*0 and final rise for the questions that met the inclusion criteria are shown in Table 7 along with the correlations for the total number of tokens from Experiment II. Only in the Contrastive Stress Condition is the correlation negative (r=-.59), while the correlation in both the Positive and Negative Affective Conditions is positive (r=.26 and r=.43, respectively). Thus, the differential relation between peak *F*0 and final rise for linguistic vs. affective distinctions is maintained for this selected subset compared to the values obtained in Experiment II, where no selection criteria were imposed.

Table 7. *Mean Correlations between Peak F0 and Final Rise for November Utterances Judged as Questions.*

	Condition		Affect		
Experiment	Neutral Stress	Contrastive Stress	Positive Affect	Negative Affect	*M*
2	.15	-.67	-.11	.52	.34
3	.25	-.59	.26	.43	.35

Discussion

The results of Experiment III show that listeners were generally able to apprehend simultaneously the affective and linguistic intentions of speakers, despite the fact that the utterances were very brief (consisting of only three syllables). With respect to the trading relation between peak *F*0 and final rise in questions, the listener judgments in this experiment provide further evidence supporting Studdert-Kennedy and Hadding-Koch's hypothesis that the trading relation is specifically linguistic in nature. The dissociation of the trading relation between questions conveying linguistic and affective contrasts found in Experiment II is confirmed when only those tokens are considered for which listeners apprehended the speakers' intent.

GENERAL DISCUSSION

The two major hypotheses of this study were supported by the results. The prediction of a trading relation between peak *F*0 and the amount of final rise in questions conveying contrastive stress was confirmed by the results of both Experiments I and II. In addition, the differential occurrence of the trading relation demonstrated in Experiment II and upheld by listeners in Experiment III suggests that increases in *F*0 associated with contrastive stress are produced differently from the increases used to convey affect.

Studies of trading relations in speech have hitherto been solely concerned with perception (e.g., Fitch, Halwes, Erickson, & Liberman, 1979; Best, Morrongiello, & Robson, 1981; Summerfield, 1975; Bailey, Summerfield, & Dorman, 1977). Although many investigators have tended to regard trading relations as a reflection of perceptual attunement to the dynamics of articulation (e.g., Repp, 1982), there has been little attempt to test this notion or develop its implications, so that the relation between perception and production has remained a matter of speculation.

Investigation of the perception-production link is particularly important because perceptual trading relations are particular to synthetically produced or altered speech, in which acoustic variables are manipulated without regard to the constraints of natural speech production. The perceptual studies by Studdert-Kennedy and Hadding (1973; Hadding-Koch & Studdert-Kennedy, 1964, 1965a,b) used hybrid stimuli formed by imposing synthetic intonation contours on a single natural utterance. Although the range of *F*0 values used to construct the contours was based on spectrographic data from natural productions (Hadding-Koch, 1961), the combination of *F*0 values in any particular stimulus was arbitrary, and all the stimuli had the same duration, to which a stylized contour was fitted. As a result, much of the variability in frequency spectrum, amplitude envelope, duration and *F*0 of naturally produced speech was missing. By contrast, the present study analyzed utterances constrained only by the scenarios devised to elicit questions with appropriate patterns of stress and affective inflections. Given the difference in experimental conditions, the closeness with which the results of the present production study match those of the perceptual studies attests to the functional reality of the trading relation between peak *F*0 and final rise. In demonstrating the trading relation in production, these results give evidence that perceptual trading relations can indeed reflect listeners' sensitivity to patterns of speech production. Moreover, this study provides specific information about the organization of *F*0 in speech, by establishing the set of circumstances under which production patterns result in a trading relation.

Two types of explanation for these findings are possible. On the one hand, both the trading relation and its limitation to a context of contrastive stress may be linguistic conventions. That is, they may be arbitrary associations, established by members of a linguistic community and learned as rules by speakers as they acquire the language (in this case, English and, presumably, Swedish). This explanation is less than satisfying, however, because it fails to address why or how the trading

relation originated. Its presence in both languages means either that it evolved in each language separately, or that English and Swedish share the trading relation as a result of their distant common ancestry in Germanic. Furthermore, this explanation requires that speakers acquire a rule for decreasing the final rise as peak *F*0 increases, only when *F*0 peaks increase to convey contrastive stress, not when they are used to convey affect. While this is not impossible, it is difficult to motivate.

On the other hand, the trading relation may reflect the articulatory dynamics of raising *F*0. Two articulatory hypotheses will be considered. Both assume that the production of the final rise is accomplished by the action of laryngeal muscles, primarily cricothyroid (CT), which controls the tension of the vocal folds, and thus *F*0 (Lieberman, Sawashima, Harris, & Gay, 1970; Atkinson, 1973). They differ on the mechanism that underlies the production of the *F*0 rise associated with stress. Lieberman (1967) proposed that the production of questions and statements in English involves the distinction between marked and unmarked breathgroups. In the unmarked breathgroup used for statements, *F*0 is strictly a function of subglottal pressure (P_S). P_S and *F*0 are typically highest at the beginning of an utterance and as P_S is used up over the course of the utterance to maintain voicing, P_S and *F*0 decrease over the duration of the breathgroup. In the marked breathgroup used for yes-no questions, a rise in *F*0 at the end of the breathgroup is effected despite falling P_S by an increase in laryngeal tension. Lieberman further proposed that a speaker effects any *F*0 rise in the middle of an utterance by increasing the flow of air through the glottis, and thus using up some of the finite subglottal-pressure reserve. In the case of a yes-no question, this loss of P_S results in a final rise that is reduced relative to what it would have been without the (previous) stress.

This hypothesis has some advantages over the "convention" explanation. For example, since it proposes an explanation based on an articulatory constraint, it does not require speakers to acquire a complex rule for when and how much to reduce the final rise. However, this proposal also has several problems. First, stress related *F*0 rises are not produced solely by using P_S, as Lieberman originally proposed. Rather, some combination of P_S and CT activity are involved (Atkinson, 1973; Gelfer, 1987).[2] Second, even if it was the case that stress production involves primarily P_S, it is unclear that the effect Lieberman suggests would occur on utterances as short as three-syllables (such as those used in this study). If it did, it would suggest a gross lack of planning in speech production; how would speakers ever get to the end of longer utterances with multiple stresses? Finally, it is not clear how Lieberman's hypothesis would account for the dissociation of the trading relation seen in Experiment II.

An alternative articulatory account of the trading relation is based on EMG studies of speech production which indicate that, in addition to P_S, CT activity is involved in the *F*0 rise associated with stress (Atkinson, 1973; Gelfer, 1987), as well as in the final *F*0 rise signaling questions (Hirano, Ohala, & Vennard, 1969; Lieberman, Sawashima, Harris, & Gay, 1970; Atkinson, 1973). Perhaps the need for CT activity on two adjacent syllables underlies the reduction in final rise following a syllable with contrastive stress. For example, based on studies of single-unit motor activity in humans, it is known that the peak effect of CT contraction on *F*0 has a latency of 70-80 ms. (Baer, 1978; 1981). Furthermore, the relaxation time of CT is longer than its contraction time (Löfqvist, Baer, McGarr, & Story, 1989). Thus, CT contractions on adjacent syllables required to produce the continuously rising question contours seen in Experiments I & II would very likely overlap in time, with the initiation of the second contraction occurring before the muscle had completely relaxed from the first contraction. This would reduce the force of the second contraction, resulting in a reduction in the amount of *F*0 rise. The lack of a trading relation for the affectively inflected questions suggests that the rise in *F*0 on the second syllable of these utterances may be produced by a mechanism different from that used for affectively neutral contrastive stress. Perhaps the most likely mechanism (especially in the case of positive affects, such as happiness) is an overall increase in P_S across the utterance, sufficient to override the effect of overlapping CT contractions.

The mechanism proposed above to account for the trading relation in intonation suggests that the production of intonational units may share certain features with the production of segmental units of speech. Because speech production is dynamic, upper-vocal tract articulatory gestures for phonetic segments often overlap temporally with the gestures for preceding or following segments. One result of this coarticulation is context-dependent variation of articulatory trajectories and ultimately of the acoustic output. Thus, for example, the direction of second and

third formant transitions for stop consonants may rise or fall depending on the relevant formant frequencies of the preceding or following vowel (Öhman, 1965). This is due to a physical (physiological) constraint on the trajectory of the tongue in moving between the positions required to produce the vowels and consonants. Öhman (1965) concluded that the neural commands for consonants and vowels in VCV sequences must be active simultaneously. The suggestion we mean to make is that an analogous phenomenon may occur in the case of intonation. That is, in the context of a preceding *F*0 prominence associated with stress, the production of final rise is altered because of a physical (physiological) constraint on the ability of the cricothyroid muscle to produce temporally overlapping gestures which raise *F*0. The result is a reduction in the magnitude of the second gesture, and therefore a reduction in the final rise.

This proposal has several advantages over both linguistic convention and Lieberman's P_S hypothesis as an explanation for the data presented in this paper. As with Lieberman's hypothesis, this proposal reduces the acquisition problem to one of discovering articulatory dynamics, for which infants seem to be quite well suited. It also eliminates the need to posit a scenario regarding how the trading relation and its specific use arose in languages as diverse as English and Swedish; presumably, these languages use the same mechanisms to raise *F*0 for stress and question final rises. This hypothesis has advantages over Lieberman's P_S hypothesis, not the least of which is that it is consistent with current data on physiological mechanisms underlying the production of stress. In addition, it offers a plausible and parsimonious explanation for both the trading relation and its specific occurrence. It may also be related to other phenomena, such as the tendency to avoid equally stressed adjacent syllables[3] and the use of rule-based substitutions in tone languages.

Although the present experiments were not specifically designed to test the question, the results are consistent with other data suggesting that speech *F*0 may be differently controlled for linguistic and affective purposes. For example, a variety of evidence suggests that the two hemispheres of the brain have somewhat different roles in the control of affective behavior both in humans and other mammalian species (Denenberg, 1981). In humans, dichotic studies with normal subjects have shown a left-ear (right-hemisphere) advantage for affective prosody (Zurif, 1974; Blumstein & Cooper, 1972; Ley & Bryden, 1982; Shipley-Brown, Dingwall, Berlin, Yeni-Komshian, & Gordon-Salant, 1988), as against the usual right-ear (left hemisphere) advantage for segmental judgments. Perceptual disorders involving emotional prosody following right-hemisphere brain injury have also been demonstrated, but deficits of linguistic prosody apparently occur following injury to either hemisphere (e.g., Tucker, Watson, & Heilman, 1977; Weintraub, Mesulam, & Kramer, 1981; Heilman, Bowers, Speedie, & Coslett, 1984; Blumstein & Cooper, 1972). Some studies of the production of linguistic intonation have found that patients with right-hemisphere damage produce normal linguistic stress and sentence intonation (Emmorey, 1987; Behrens, 1988), but others have reported deficits after right hemisphere damage (Shapiro & Danly, 1985; Weintraub, Mesulam, & Kramer, 1981). Although far from unanimous in their indications, these studies point to a different degree of involvement of the left and right hemispheres in linguistic and affective prosody, including intonation. The results of the present study are consistent with the hypothesis that the control of linguistic intonation is functionally separable from the affective use of *F*0.

REFERENCES

Atkinson, J. E. (1973). *Aspects of intonation in speech: Implications from an experimental study of fundamental frequency*. Unpublished doctoral dissertation. University of Connecticut.

Baer, T. (1978). The effect of single-motor-unit firings on fundamental frequency. *Journal of the Acoustical Society of America, 64*, S90(A).

Baer, T. (1981). Investigation of the phonatory mechanism. ASHA Reports (11): Proceedings of the Conference on the Assessment of Vocal Fold Pathology. Hart, MO: Ludlow.

Bailey, P., Summerfield, Q., & Dorman, M. (1977). On the identification of sine-wave analogues of certain speech sounds. *Haskins Status Reports, SR 51/52*, 1-25.

Behrens, S. (1988). The role of the right hemisphere in the production of linguistic stress. *Brain and Language, 33*, 104-127.

Best, C., Morrongiello, B., & Robson, R. (1981). The perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics, 29*, 191-211.

Blumstein, S., & Cooper, W. (1972). Hemispheric processing of intonation contours. *Cortex, 10*, 146-158.

Blumstein, S., & Lieberman, P. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.

Bolinger, D. L. (1958). A theory of pitch accent in English. *Word, 14*, 109-149.

Bolinger, D. L. (1978). Intonation across languages. In J. P. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language. Volume 2: Phonology*. Stanford: Stanford University Press.

Cohen, A. (1972). Some observations on the pitch of questions. In Valdman, A. (Ed.), *Papers in memory of Pierre Delattre*. London: Mouton.

Cohen, A., Collier, R., & t'Hart, J. (1982). Declination: Construct or intrinsic feature of speech pitch? *Phonetica, 39*, 254-273.

Cooper, W., & Sorenson, J. (1981). *Fundamental frequency in sentence production*. New York: Springer-Verlag.

Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 864-881.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Davitz, J. R. (1964). A review of research concerned with facial and vocal expressions of emotion. In J. R. Davitz (Ed.), *The communication of emotional meaning*. New York: McGraw-Hill.

Denenberg, V. (1981). Hemispheric laterality in animals and the effects of early experience. *Behavioral and Brain Sciences, 4*, 1-49.

Emmorey, K. (1987). The neurological substrates for prosodic aspects of speech. *Brain and Language, 30*, 305-320.

Fairbanks, G., & Provonost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotions. *Speech Monographs, 6*, 87-104.

Fairbanks, G., & Hoaglin, L. (1941). An experimental study of the duration characteristics of the voice during the expression of emotions. *Speech Monographs, 8*, 85-90.

Ferguson, G. (1981). *Statistical analysis in psychology and education*. New York: McGraw-Hill Book Company.

Fitch, H., Halwes, T., Erickson, D., & Liberman, A. M. (1979). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception and Psychophysics, 27*, 343-350.

Fries, C. C. (1964). On the intonation of yes-no questions in English. In D. Abercrombie (Ed.), *In honour of Daniel Jones*. London: Longmans.

Fry, D. B. (1955). Duration and intensity as acoustic correlates of linguistic stress. *Journal of the Acoustical Society of America, 35*, 765-769.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1*, 126-152.

Gelfer, C. (1987). *A simultaneous physiological and acoustic study of fundamental frequency*. Unpublished doctoral dissertation, City University of New York.

Hadding-Koch, K. (1961). *Acoustico-phonetic studies of intonation in southern Swedish*. Lund: Geerups.

Hadding-Koch, K., & Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. *Phonetica, 11*, 175-185.

Hadding-Koch, K., & Studdert-Kennedy, M. (1965a). A study of semantic and psychophysical test responses to controlled variations in fundamental frequency. *Studia Linguistica*, 17, 65-76.

Hadding-Koch, K., & Studdert-Kennedy, M. (1965b). Intonation contours evaluated by American and Swedish test subjects. *Proceedings of the Vth International Congress of Phonetic Sciences*, Munster, 1964. Basel: S. Karger.

Heilman, K., Bowers, D., Speedie, L., & Coslett, H. (1984). Comprehension of affective and non-affective prosody. *Neurology, 34*, 917-921.

Hirano, M., Ohala, J., & Vennard, W. (1969). The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. *Journal of Speech and Hearing Research, 12*, 616-628.

Jassem, W. (1972). The question-phrase fall-rise in British English. In A. Valdman (Ed.), *Papers in Memory of Pierre Delattre*. Paris: Mouton.

Kramer, E. (1964). Elimination of verbal cues in judgments of emotion from voice. *Journal of Abnormal and Social Psychology, 68 (4)*, 390-396.

Lee, W. R. (1956). English intonation: A new approach. *Lingua, 4*, 361.

Lee, W. R. (1960). An English intonation reader. Macmillan: London.

Lee, W. R. (1980). A point about the rise-endings and fall-endings of yes-no questions. In L. R. Waugh, & C. H. van Schooneveld (Eds.), *The melody of language intonation and prosody*. Baltimore: University Park Press.

LeGuinn, U. (1968). *A wizard of earthsea*. Boston: Houghton Mifflin Co.

Ley, R., & Bryden, M. (1982). A dissociation of right and left hemisphere effects for recognizing emotional tone and verbal content. *Brain and Cognition, 1*, 3-9.

Lieberman, P. (1967). *Intonation, perception and language*. Cambridge, MA: MITT Press.

Lieberman, P., Sawashima, M., Harris, K., & Gay, T. (1970). The articulatory implementation of the breath-group and prominence: Cricothyroid muscular activity in intonation. *Language, 46*, 312-327.

Löfqvist, A., Baer, T., McGarr, N., & Story, R. (1989). The cricothyroid muscle and voicing control. *Journal of the Acoustical Society of America, 85*, 1314-1321.

McCluskey, K., Albas, D., Niemi, R., Cuevas, C., & Ferrer, C. (1975). Cross-cultural difference in the perception of the emotional content of speech: A study of the development of sensitivity in Canadian and Mexican children. *Developmental Psychology, 11 (5)*, 551-555.

Ohman, S. E. G. (1965). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39*, 151-168.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America, 66*, 363-369.

Pike, K. (1945). *The intonation of American English*. University of Michigan Press: Ann Arbor.

Repp, B. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*, 81-110.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*, 143-165.

Scherer. K. R., Ladd, D. R., & Silverman, K. E. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America, 76*, 1346-1356.

Shapiro, B., & Danly, M. (1985). The role of the right hemisphere in the control of speech prosody in propositional and affective contexts. *Brain and Language, 25*, 19-36.

Shipley-Brown, F., Dingwall, W., Berlin, C., Yeni-Komshian, G., & Gordon-Salant, S. (1988). Hemispheric processing of affective and linguistic intonation contours in normal subjects. *Brain and Language, 33*, 16-26.

Signal Technology, Inc. (1978). *Interactive Laboratory System*. Goleta, CA: Signal Technology, Inc.

Starkweather, J. A. (1961). Vocal communication of personality and human feelings. *Journal of Communication, 11*, 63-72.

Studdert-Kennedy, M., & Hadding, K. (1973). Auditory and linguistic processes in the perception of intonation contours. *Language and Speech, 16*, 293-313.

Summerfield, Q. (1975). *Information-processing analyses of perceptual adjustments to source and context variables in speech*. Unpublished doctoral dissertation, Queen's University of Belfast.

Tartter, V. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics, 27*, 24-27.

Tucker, D., Watson, R., & Heilman, K. (1977). Discrimination and evocation of affectively intoned speech in patients with right parietal disease. *Neurology, 27*, 947-950.

Uldall, E. T. (1960). Attitudinal meanings conveyed by intonation contours. *Language and Speech, 3*, 223-234.

Ultan, R. (1978). Some general characteristics of interrogative systems. In J. P. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language. Volume 2: Phonology*. Stanford: Stanford University Press.

Weintraub, S., Mesulam, M., & Kramer, L. (1981). Disturbances in prosody—A right hemisphere contribution to language. *Archives of Neurology, 38,* 742-744.

Williams, C. E., & Stevens, K. N. (1969). On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Medicine, 40,* 1369-1372.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustic correlates. *Journal of the Acoustical Society of America, 52,* 1238-1250.

Williams, C. E., & Stevens, K. N. (1981). Vocal correlates of emotional states. In J. K. Darby, (Ed.), *Speech Evaluation in Psychiatry.* Grune and Stratton: London.

Zurif, E. (1974). Auditory lateralization: Prosodic and syntactic factors. *Brain and Language, 1,* 91-404.

FOOTNOTES

Perception & Psychophysics, 57(2), 159-174 (1995).

†Now at the Department of Psychology, Stanford, University.

‡Also University of Connecticut, Storrs.

[1]Indeed, it might be argued that any use of final rise to convey questions reflects the emotional state of the speaker, since in most languages syntactic devices are available which can be used to indicate questions without the use of final rise. For example, question words, tags, word order, and particles are among the structures used to express interrogation in many of the world's languages, including English (Ultan, 1978; Bolinger, 1978). Alternatively, it may be that the use of final rise was, at some stage of prehistory, an unequivocal indication of the affective state of the speaker, but has undergone historical change, such that it is now used as a formal device, an alternative to syntactic devices, such as wh-movement, without necessarily implicating the speaker's emotional state.

[2]Lieberman seems to have accepted some role for CT in stress-related *F*0 rises in his recent book with Blumstein (Blumstein & Lieberman, 1988).

[3]We do not mean to suggest that this is the only reason for avoiding equally stressed adjacent syllables. Metrical factors are also likely to be relevant.

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 133-140

Orthographic Representation and Phonemic Segmentation in Skilled Readers: A Cross-language Comparison*

Ilana Ben-Dror,† Ram Frost,‡ and Shlomo Bentin†, ‡

The long lasting effect of reading experience in Hebrew and English on phonemic segmentation was examined in skilled readers. Hebrew and English orthographies differ in the way they represent phonological information. Whereas each phoneme in English is represented by a discrete letter, in unpointed Hebrew most of the vowel information is not conveyed by the print and, therefore, a letter often corresponds to a CV utterance (i.e., a consonant followed by a vowel). Adult native speakers of Hebrew or English, presented with words consisting of a consonant, a vowel and then another consonant were required to delete the first "sound" of each word and to pronounce the remaining utterance as fast as possible. Hebrew speakers deleted the initial CV segment instead of the initial consonant more often than English speakers, for both Hebrew and English words. Moreover, Hebrew speakers were significantly slower than English in correctly deleting the initial phoneme, and faster in deleting the whole syllable. These results suggest that the manner in which orthography represents phonology not only affects phonological awareness during reading acquisition, but also has a long-lasting effect on skilled readers' intuitions concerning the phonological structure of their spoken language.

Phonological awareness is the ability to recognize and manipulate internal phonemic constituents of spoken words. Previous research has provided ample evidence that this ability is necessary for reading acquisition and is related to skilled reading performance (for recent reviews see Bentin, 1992; Goswami & Bryant, 1990). For example, reliable correlations were found between children's ability to manipulate subword units and the rate and efficiency of learning to read (Goswami & Bryant, 1990; Liberman, Shankweiler, Liberman, Fowler, & Fisher, 1977; Mann & Liberman, 1984; Treiman, 1985). In addition, phonological awareness in kindergarten was found to be a good predictor of reading success in the early school years (Bradley, 1989; Bradley & Bryant, 1983; Lundberg, Olofsson, & Wall, 1980; Mann, 1984; Stanovich, Cunningham, & Cramer, 1984).

This study was supported by a grant from the Israeli Foundation Trustees and partly by National Institute of Child Health and Human Development Grant HD 01994 to Haskins Laboratories. Ilana Ben Dror was supported by post-doctoral stipends from the Lady Davis and Golda Meir Foundations.

A causal connection between phonological skills and reading acquisition has been supported by studies showing that intervention aimed at improving phonological skills facilitates reading acquisition (Ball & Blachman, 1988, 1991; Bentin & Leshem, 1993; Blachman, 1989; Bradley & Bryant, 1983; Lundberg, Frost, & Peterson, 1988). Other studies have shown, however, an inverse causal connection, that is, that exposure to literacy enhances phonological awar81990; Morais, Bertelson, Cary, & Alegria, 1986). Together, these results suggest a strong bidirectional influence between reading acquisition and phonological awareness. Probably, the exposure to clearly defined orthographic segments triggers awareness of coarticulated phonemic segments, while at the same time this awareness fosters the acquisition of grapheme-to phoneme correspondence rules. This interpretation is further supported by comparing the effects of different reading instruction methods on reading skills. Alegria, Pignot and Morais (1982) reported that children who learned to read by analytic methods emphasizing letter-sound correspondences, performed better on tests of

phonemic segmentation than children who learned by holistic methods.

These results suggest that the manner in which the writing system represents the spoken language may influence phonological awareness. Support for this claim is gained from cross-linguistic studies. Mann (1986) compared phonological awareness of syllables and phonemes in Japanese and American first graders and found that Japanese children performed more poorly than American children on tests assessing awareness of phonemes but not of syllables. Mann argued that Japanese children's performance was influenced by their reading experience in a syllabary orthography, whereas American children were affected by their reading experience in an alphabetic orthography. Her conclusions are supported by Read, Zhang, Nie, and Ding (1986) who showed that literate Chinese adults who learned to read the alphabetic (pinyin) orthographic system performed better in phonemic segmentation tests than literate adults who read only the logographic (kanji) system.

A possible conclusion from these studies is that the size of the phonological unit that the beginning reader becomes aware of is affected by the size of the speech segment into which orthographic units are mapped. In the present study, we investigated this hypothesis by comparing phonological sensitivity of skilled adult readers trained initially to read either Hebrew or English. We sought to examine the influence of different orthography-to-phonology mapping rules on the ability of *mature* readers to manipulate the various segments of the spoken words.

The Hebrew writing system is characterized by several properties that make it interesting for comparison with other alphabetic orthographies such as English orthography (e.g., Frost & Bentin, 1992). In Hebrew, letters represent mostly consonants while most of the vowels are represented by diacritic marks (dots and dashes). Some vowels however are represented by letters (which have dual function and represent either a consonant or a vowel). There are two modes of writing Hebrew: pointed and unpointed. The pointed writing system contains all the diacritical marks and is used mainly for children books, holy scripts and poetry. The unpointed print uses the same letter characters as the pointed system (including the vowel letters) but omits the diacritic marks. The pointed system is initially taught in the early elementary grades. However, starting in the third grade, the vowel marks are gradually omitted from textbooks, and adult readers use the unpointed writing system almost exclusively. Hence, although Hebrew has an alphabetic orthography, its basic orthographic units usually represent more than single phonemes. Because the letters are mostly consonants onto which vowels are *subsequently* attached, these orthographic units usually correspond to consonant-plus-vowel (CV) utterances.

In the present study, we examined whether the size of the unit represented by the orthography affects the size of the segments that mature readers are aware of. Specifically, we hypothesized that reading in Hebrew, where alphabetic units represent mostly CV segments, should foster awareness of spoken word segments of that size (CV). In contrast, reading in English, in which most letters are mapped into phonemic segments, should enhance awareness of single phonemes. We were interested in the intuition of adult readers concerning the phonemic structure of their spoken language, as well as in their ability to manipulate single phonemes.

Several studies have suggested that for skilled readers, orthographic and phonological representations interact so that orthographic knowledge affects the recognition of words in the auditory modality. It has been shown that lexical decisions to spoken words are facilitated if successive words share the same spelling (Jakimik, Cole, & Rudnicky, 1980). Similarly, using the naming task, Tanenhaus, Flanigan, and Seidenberg (1980) have demonstrated a visual-auditory interference in a Stroop paradigm. However, although it seems evident that reading and listening could share one lexicon, allowing identical messages to be understood in the two modalities in the same way, it is possible that phonological awareness and the basic phonological skills of adult native speakers are independent of the special characteristics of their writing system.

If orthographic knowledge affects phonological skills, such as phonemic segmentation, then English and Hebrew speakers should differ, for example, in their ability to omit the first consonant of a spoken word and pronounce the remaining phonemes (a phoneme deletion task). Whereas this task should be fairly simple for readers of English, the reading experience of Hebrew readers would cause them to delete the initial CV units of Hebrew words. Moreover, even a correct deletion of the initial phoneme instead of the initial CV unit would involve greater cognitive effort for Hebrew readers and, consequently,

would result in slower deletion latencies for Hebrew than for English speakers. In contrast, similar performance of the two subject groups would support a view that basic phonological skills of literate adults are not affected by their reading experience. A second aim of the present study was to examine the effect a first language may have on phonemic segmentation in a second language, that is, how Hebrew bilingual readers segment words in English, and vice versa.

Method

Subjects. Fifty two subjects participated in the experiment for course credit or for payment. Twenty six subjects were bilingual native speakers of Hebrew from The Hebrew University, and 26 subjects were bilingual speakers of English from a 1- year Hebrew program for overseas students.

Tests and Materials. Phonemic sensitivity was measured by a phoneme deletion test, a task that is commonly used to assess phonological awareness. There were two word lists, one containing English words and one containing Hebrew words. Each list contained 28 monosyllabic items consisting of a consonant, then a vowel and a final consonant (CVC). Half of the items in each word list were words that have identical phonological structures in the two languages (but obviously differ in the pronunciation of their surface phonetics). For example, English words such as *but* and *gun* are homophonic with Hebrew words that have the same phonemic sequence but have, naturally, different meaning (/bʌt/ meaning "daughter" and /gʌn/ meaning "garden"). However, whereas in English the vowel of each of these words is represented by an independent letter, in Hebrew the vowels are omitted in print (e.g., BT for /bʌt/, and GN for /gʌn/). Comparing performance on deletion tests for words that "sound the same" across languages, may provide strong evidence concerning the influence of orthographic representation on phonemic awareness. The other half of the items in the lists were pairs of CVC Hebrew and English words that were matched on the first and second phonemes. The distribution of types of final consonants was similar across languages.

In addition to the above between-language comparison, the study included a within-language comparison of performances for two types of words within the Hebrew orthography. Fourteen of the 28 Hebrew stimuli were words that are written with a vowel letter (i.e., the vowel is explicitly represented in print as in the word KIR, (meaning wall) in which the vowel /i/ is represented by the Hebrew letter "I"). The other fourteen words were words in which the vowels are not represented in print (such as BT or GN). Comparing phonemic deletion performance, within the same language for words that have the same internal phonological structure (CVC) but differ in their orthographic representation may provide further insight into the influence of orthographic representations on phonemic segmentation.

Procedure and apparatus. Each subject was tested individually. Half of the subjects in each native language group heard the English words first and the Hebrew words second. The other half heard the Hebrew words first and the English second. Subjects were told that they were about to hear some words in Hebrew (or in English) and were instructed to delete the first "sound" of each word and say as fast as possible "what is left of it". The first "sound" was not specified by the experimenter, and no feedback was given to the subjects following their responses. This allowed *intuitive* judgments about the size of the phonemic segment that subjects deleted. Reaction time (RT) was measured from the onset of the experimenter's uttered stimulus (in each trial) to the onset of the subject's response using a voice key. The voice key was attached to an electronic counter-timer and RTs were logged manually by the experimenter along with the subject's response.

Results

"Correct" response in the present study was considered to be a VC utterance reflecting deletion of the initial consonant only (e.g., /ʌt/ following the word BUT). Errors consisted mainly of omitting the CV segment, thereby producing the final consonant. For each word in each language, the percentage of correct responses and the mean reaction time (RT) were calculated separately for Hebrew and English native speakers. These data were further categorized by the order of presentation (i.e., whether the subject was presented with Hebrew followed by English words or vice versa). The RT analysis was based on correct trials only (see Table 1).

These data were analyzed by three-factor analysis of variance (ANOVA) in an item analysis.[1] The effect of the language of the stimulus was averaged within subjects groups, and the effects of the native language of the subjects and the order of presentation were assessed within items.

Table 1. *Reaction time (in milliseconds) and percentage of correct responses for deletion of the first phoneme in Hebrew and English words.*

Measure of performance	Native Hebrew speakers		Native English speakers	
	Presented first	Presented second	Presented first	Presented second
		Hebrew Words		
Reaction time	1,149 (41)	904 (34)	736 (11)	678 (9)
Percentage correct	63 (2.3)	76 (1.3)	92 (0.7)	100 (0.0)
		English Words		
Reaction time	933 (25)	1,040 (25)	742 (18)	664 (14)
Percentage correct	79 (1.4)	74 (0.9)	100 (0.0)	92 (0.0)

Note. Numbers in parentheses are standard errors of the means. Reaction times are for correct trials only.

The analysis of the accuracy data showed that the percentage of correct responses was higher for English words (86.25%) than for Hebrew words (82.75%), $F(1,54) = 16.53$, $MS_e = 43.2$, $p < .001$; that native English speakers were more accurate than native Hebrew speakers (97.5% and 73.2%, respectively), $F(1,54) = 678.7$, $MS_e = 42.9$, $p < .001$; and that the percentage of correct responses was higher for the second test (85.5%) than for the first test (83.5%), $F(1,54) = 8.82$, $MS_e = 26.9$, $p < .005$. Both the effect of native language of the subject and the effect of the order of presentation interacted with the effect of the stimulus language, $F(1,54) = 14.19$, $MS_e = 42.9$, $p < .001$; and, $F[1,54) = 155.6$, $MS_e = 26.9$, $p < .001$; respectively. Finally, a significant interaction of order of presentation and native language of the subject suggested that the order of presentation affected the performance of native Hebrew speakers more than that of native English speakers, $F(1,54) = 5.82$, $MS_e = 30.7$, $p < .025$. The three-way interaction was not significant, $F(1,54) < 1.0$. Because of insufficient variation in the accuracy scores of the native English speakers (i.e., many subjects having zero errors), the nature of the interaction could not be reliably investigated any further in this group. For native Hebrew speakers, a two-way ANOVA showed that the order of presentation influenced performance with Hebrew and English words, but in opposite directions, $F(1,54) = 44.8$, $MS_e = 54.5$, $p < .001$. For the Hebrew words accuracy was lower when they were presented first than when they were presented second, $t(55) = 2.058$ $p < .05$]; for English words the accuracy was not significantly different regardless of whether they were presented before or after the Hebrew words, $t(55) = 0.244$.

The analysis of RTs showed that overall, responses were equally fast for English (845 ms) and Hebrew (867 ms) words, $F(1,54) = 0.965$. However, native English speakers were much faster to respond (705 ms) than native Hebrew speakers (1006 ms), $F(1,54) = 338.63$, $MS_e = 15055$, $p < .001$], regardless of the language of the materials. Overall, responses were slower in the first test (890 ms) than in the second test (821 ms) $F(1,54) = 22.93$, $MS_e = 11429$, $p < .001$]. But order of testing interacted with the stimulus language $F(1,54) = 33.64$, $MS_e = 11429$, $p < .001$]. Furthermore, a significant three-way interaction suggested that the interaction between the effect of word language and the effect of the order of presentation was different for native Hebrew and native English speakers $F(1,54) = 43.36$, $MS_e = 11064$, $p < .001$.

To clarify the source of the three-way interaction, we conducted two-way ANOVAs for Hebrew and English speakers separately. These analyses revealed that order of presentation and stimulus-language interacted for native Hebrew speakers, $F(1,54) = 44.3$, $MS_e = 19452$, $p < .001$, but not for the native English speakers, $F(1,54) < 1.0$]. Hebrew speakers responded to Hebrew words more slowly when they were presented first than when they were presented following the English materials, $t(55) = 1.955$, $p < .06$. In contrast, they responded faster to English words when they were tested first than when they followed the Hebrew words, $t(55)=6.5$, $p < .001$.

Table 2 presents subjects' performance with the subset of stimuli that were phonologically identical in Hebrew and English. The pattern of results for this subset of words is very similar to the pattern found for the entire set. That is, native Hebrew and English speakers differed in their performance in segmenting words that are pronounced similarly in the two languages.

Table 2 *Reaction time (in milliseconds) and percentage of correct responses for deletion of the first phoneme in phonologically identical Hebrew and English words.*

Measure of performance	Native Hebrew speakers		Native English speakers	
	Presented first	Presented second	Presented first	Presented second
		Hebrew Words		
Reaction time	1,272 (55)	875 (28)	730 (17)	668 (12)
Percentage correct	37 (3.7)	76 (2.6)	92 (0.8)	100 (0.0)
		English Words		
Reaction time	942 (33)	1,100 (32)	769 (23)	684 (16)
Percentage correct	76 (2.2)	73 (1.0)	100 (0.0)	92 (0.0)

Note. Numbers in parentheses are standard errors of the means. Reaction times are for correct trials only.

Finally, we compared performance for Hebrew words in which the vowel is represented in print only by dots and words in which the vowel is represented by a letter (Table 3).

Table 3. *Reaction time (in milliseconds) and percentage of correct responses for deletion of the first phoneme in Hebrew words that do and do not include printed vowels.*

Measure of performance	Native Hebrew speakers	Native English speakers
	Words with vowels	
Reaction time	985 (45)	715 (11)
Percentage correct	84 (1.0)	96 (0.5)
	Words without vowels	
Reaction time	1,074 (34)	699 (12)
Percentage correct	65 (2.0)	96 (0.4)

Note. Numbers in parentheses are standard errors of the means, Reaction times are for correct trials only.

ANOVA showed that the effect of word type on performance differed for Hebrew and English native speakers. For Hebrew speakers, the percentage of correct deletions of initial phonemes in CVC words was higher when the vowel was represented in print by a letter than when it was represented only by points, and RTs to the correctly segmented words were faster if the vowel was represented by a letter than if it was not. In contrast, for English speakers, the percentage of correct deletions was not influenced by the word type, and RTs for words with vowel letters were even slower than RTs for words without vowel letters. The interaction of word type by subjects' native language was significant for both accuracy and speed of responses, $F(1,26) = 18.18$, $MS_e = 18.9$, $p < .001$ and $F(1,26)=4.48$, $MS_e = 8732$, $p < .05$, respectively.

Before drawing firm conclusions from these results, we had to examine whether the poorer performance of Hebrew than of English speakers does not stem from simple group differences in verbal skills, which are related to differences in academic background. For example, Treiman, Fowler, Gross, Berch, and Weatherston (in press) have shown that performance in the phoneme deletion task is correlated with university selectivity. To address this methodological issue, Hebrew and English native speakers were tested in a syllable deletion task. Using an identical method and apparatus, 24 Hebrew and 24 English speakers were presented with 24 Hebrew and English bisyllabic words and were explicitly required to omit their initial syllables instead of their initial "sounds." The results are presented in Table 4.

Table 4. *Reaction time (in milliseconds) and percentage of correct responses for deletion of the first syllable in Hebrew and English words.*

Measure of performance	Native Hebrew speakers	Native English speakers
	Hebrew words	
Reaction time	790 (15.9)	912 (15.5)
Percentage correct	98.5	97.8
	English words	
Reaction time	846 (15.9)	915 (14.7)
Percentage correct	97	94

Note. Numbers in parentheses are standard errors of the means. Reaction times are for correct trials only.

In contrast to results for the deletion of initial phonemes, Hebrew speakers were overall faster than English speakers in deleting the initial syllable for both Hebrew and English materials, $F1(1,46) = 141$, $MS_e = 1551$, $p < 0.001$. There was no significant difference in RT between omitting the first syllable of Hebrew and English words, $F1(1,46) = 2.0$, MS_e 0, $p < 0.15$. The interaction between speakers and material was significant: Native Hebrew speakers were faster in deleting the first syllable of Hebrew than of English words, whereas language did not affect the Native English speakers $F1(1,46) = 10.7$, $MS_e = 1551$, $p < 0.002$. The small percentage of errors did not allow a reliable statistical analysis. Nonetheless, it is evident that in contrast to the phoneme deletion task, Hebrew speakers were not less accurate than English speakers in syllable deletion. In fact, most errors were made by English speakers for English words, and consisted of omitting the initial phoneme rather than the first syllable.

DISCUSSION

The results of the present study indicate that the phonological sensitivity of fluent readers and their ability to manipulate phonemic segments may be influenced by the way phonological information is represented by the orthography. When asked to remove the first sound of English and Hebrew words composed of CVC trigrams, native Hebrew speakers tended to remove the initial CV combinations rather than the initial consonants more often than native English speakers. Two findings reveal that this difference reflects not merely different understandings of the deletion test (e.g., confusing the removal of the first "sound" with the removal of the first letter) but a genuine cognitive difference in manipulating phonemes. First, native Hebrew speakers made more errors in both languages, suggesting that they did not just omit the first letter and produce the remaining utterance. Second, and more important, they were overall much slower than native English speakers in *correctly* deleting only the initial phoneme for both Hebrew and English materials. This outcome suggests a genuinely greater difficulty in detaching single phonemes in the phoneme deletion task. The results from the syllable deletion task strongly reinforce this conclusion. When it was not the initial phoneme that had to be detached, but the initial syllable, Hebrew readers performed significantly better than English readers.

Within-language-group comparisons revealed a marked difference in the way Hebrew and English speakers were affected by the language of the presented stimulus. Whereas English speakers performed almost identically with Hebrew and English materials, Hebrew speakers tended to isolate CV segments more often with Hebrew than with English words, even when the words had identical phonemic sequences in the two languages. Hebrew speakers also differed from English speakers in their susceptibility to the order of material presentation. The RTs analyses suggest that the order of presentation reflects mostly a practice effect for English speakers, who were faster in the second test block than in the first, regardless of the stimulus language. In contrast, performance of Hebrew speakers depended heavily on the language of their initial test session. An initial exposure to English materials had a beneficial effect on Hebrew speakers' subsequent performance with Hebrew words; RTs to Hebrew words were 245 ms faster if the first testing session was conducted in English. A similar pattern was revealed in the accuracy scores. However, an opposite effect was found with English materials: An initial exposure to Hebrew words hindered performance in the subsequent presentation of English words. Thus, Hebrew speakers were some 100 ms slower to detach the first phoneme of English words if they were first tested with Hebrew materials.

This result provides some insight concerning the cognitive procedures used by the subjects in the phoneme deletion task. When asked to delete the first sound of the word, subjects probably invoked the word's orthographic representation and based their decision, in part, on the deletion of the first letter. Thus, because in all of the English words that we employed, the second letter was a vowel, Hebrew subjects probably correctly realized that what should be deleted was the initial consonant and generalized this strategy to the Hebrew words as well. An opposite effect occurred when Hebrew words were presented first; in this case, the previous exposure to Hebrew materials was detrimental to segmenting the English words correctly. This interpretation is supported by the performance with Hebrew words that contain vowel letters. Hebrew speakers deleted the first phoneme of these words more often and much faster than words that do not include a vowel letter. This within-language effect reflects a strategy of invoking an orthographic representation in the phoneme deletion task.

The results, however, cannot be explained only in terms of invoking an orthographic representation in the task. First, performance of English

speakers was similar across sessions and across stimuli, and did not reflect a fine-tuning to the different orthographic structures of the spoken word in the two languages. More important, regardless of the order of presentation and the materials to be segmented, Hebrew speakers were always slower than English speakers in the phoneme deletion task, even when they performed the task correctly. Therefore, we suggest that the different trends observed in the two groups' abilities to delete the first consonants of words, represents a cognitive difference in manipulating phonemes. This cognitive difference is probably influenced by the basic phonemic awareness developed early in life and modulated by the nature of the grapheme-to-phoneme rules specific to the languages' orthography. This view is consistent with the results obtained in the syllable deletion test. Because the Hebrew unpointed graphemes often represent syllables, Hebrew speakers performed better than English speakers in this test, in sharp contrast to their performance in phoneme deletion.

The results of the present study may clarify the mechanism by which reading acquisition fosters the development of phonological skills, and the ability to segment words into their phonemic constituents during the early school years. In contrast to speech, in which the phonemes are coarticulated and cannot be easily disentangled from one another, in writing the phonemes are represented by discrete units—the letters. When children are aware of the fact that words are composed of smaller meaningless units (i.e., have basic phonological awareness), exposure to an alphabet may help in determining the size of these units. In an alphabetic orthography, letters are mapped onto single phonemes, and therefore exposure to the alphabetic principle may help children realize that the smallest phonological unit is the phoneme. A writing system in which letters represent single phonemes has apparently long lasting effects that extend to adult readers as well as children. When the native English speakers in our study were asked to delete the first sound in words, they isolated the initial sound most of the time, in English as in Hebrew. They did this because in English letters always represent single phonemes. In Hebrew, in contrast, although letters denote mainly single consonants, these consonants are combined and pronounced with the following vowels because the orthographic symbols that denote vowels are usually absent. Therefore, the basic subword phonological unit induced by exposure to Hebrew letters may take the form of a CV phonological unit. This may have caused the enhanced tendency of the native Hebrew speakers in the present study to isolate CV units in the deletion task both in Hebrew and in English.

In summary, our results support the claim that the way in which orthography represents phonology affects phonological awareness (Mann, 1986; Read et al., 1986). However, the present study suggests that this effect is not restricted to the phase of reading acquisition. Rather, it has a long lasting influence on skilled readers intuition concerning the phonological structure of their spoken language, and even on their basic phonological skills.

REFERENCES

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. *Memory and Cognition, 10,* 451-556.

Ball, E. & Blachman, B. (1988). Phoneme segmentation training: Effect on reading readiness. *Annals of Dyslexia, 38,* 208-225.

Ball, E., & Blachman, B. (1991). Does phoneme awareness in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26(1),* 49-66.

Bentin, S. (1992). Phonological awareness, reading, and reading acquisition: A survey and appraisal of current knowledge. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67-84). North-Holland: Elsevier.

Bentin, S., Hammer, R., & Cahan, S. (1991). The effects of aging and first year schooling on the development of phonological awareness. *Psychological Science, 2(4),* 271-274.

Bentin, S., & Leshem, H. (1993). On the interaction of phonologic awareness and reading acquisition: it's a two-way street. *Annals of Dyslexia.*

Blachman, B. (1989). Phonologic awareness and word recognition: Assessment and intervention. In A. G. Kamhi & H. W. Catts (Eds.), *Reading disabilities: A developmental language perspective* (pp. 133-158). Boston: College Hill Press.

Bradley, L. (1989). Predicting learning disabilities. In J. J. Dumont & H. Nakken (Eds.), *Learning disabilities: Cognitive, social and remedial aspects* (pp. 1-17). Amsterdam: Swets & Zeitlinger.

Bradley, L. & Bryant, P. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301,* 419-421.

Frost, R., & Bentin, S. (1992). Reading consonants and guessing vowels: Visual word recognition in Hebrew orthography. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 27-44). North-Holland: Elsevier.

Goswami, U., Bryant, P. (1990). *Phonologic skills and learning to read.* East Sussex: Erlbaum.

Jakimik, J, Cole, R. A., & Rudnicky, A. I. (1980). *The influence of spelling on speech perception.* Paper presented at the XXI annual meeting of the Psychonomic Society. St Louis, Missouri.

Liberman, I. Y., Shankweiler , D., Liberman, A. M., Fowler, C., & Fisher, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), *Towards a psychology of reading* (pp. 207-226). Hillsdale, NJ: Erlbaum.

Lundberg, I., Frost, J., & Peterson, O. P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263-284.

Lundberg, I., Olofsson, A., & Wall. S. (1980). Reading and spelling skills in the first school years predicted from phonemic

awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

Mann, V. A. (1984). Longitudinal prediction and prevention of early reading difficulty. *Annals of Dyslexia, 34*, 117-136.

Mann. V. A. (1986). Phonological awareness: The role of reading experience. *Cognition, 24*, 69-52.

Mann, V. A., & Liberman, I. Y. (1984). Phonologic awareness and verbal short term memory. *Journal of Learning Disabilities*, 592-598.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24*, 45-64.

Read, C. A., Zhang, Y., Nie, H., Ding, B. (1986). The ability to manipulate speech sounds depends on knowing the alphabetic reading. *Cognition, 24*, 31-44.

Stanovich. K. E., Cunningham, A., & Cramer, B. (1984). Assessing Phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 3*, 175-190.

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition, 8*, 513-520.

Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology, 39*, 161-181.

Treiman, R., Fowler, C. A., Gross, J., Berch, D., Weatherston, S. (1995). Syllable structure or word structure: Evidence for onset and rime with dysyllabic and trisyllabic stimuli. *Journal of Memory and Language, 34*, 132-155.

FOOTNOTES

*Also appears in *Psychological Science*, 1-6 (1995).

†School of Education, The Hebrew University.

‡Department of Psychology, The Hebrew University.

[1] A subject analysis could not be carried out because the factors of order of presentation and stimulus language were not independent across subjects (subjects who were presented with Hebrew stimuli first, were necessarily presented with English stimuli second, and vice versa). Therefore, the variance in each level of order of presentation is not independently distributed for stimulus language and speakers' language, across subjects.

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 141-160

Expressive Timing in Schumann's "Träumerei": An Analysis of Performances by Graduate Student Pianists*

Bruno H. Repp

Statistical analyses were conducted on the expressive timing patterns of performances of Schumann's "Träumerei" by 10 graduate student pianists who played from the score on a Yamaha Disclavier after a brief rehearsal. A previous study of acoustic recordings of "Träumerei" by 24 famous pianists [B. H. Repp, J. Acoust. Soc. Am. *92*, 2546-2568 (1992)] provided "expert" timing data for comparison. In terms of group average timing pattern, individual shaping of *ritardandi*, and within-performance consistency, the students turned out to be quite comparable to the experts. This demonstrates that precision in expressive timing does not require extensive study and practice of the music at hand, only general musical and technical competence. Subsequent principal components analyses on the students' timing patterns revealed that they were much more homogeneous than the experts'. Individual differences among student pianists seemed to represent mainly variations around a common performance standard (the first principal component), whereas expert performances exhibited a variety of underlying timing patterns, especially at a detailed level of analysis. Experienced concert artists evidently feel less constrained by a performance norm, which makes their performances more interesting and original, hence less typical. Since the norm may represent the most natural or prototypical timing pattern, relatively spontaneous performances by young professionals may be a better starting point for modelling expressive timing than distinguished artists' performances.

INTRODUCTION

Musical performance is one of the most intricate and highly developed skills humans are capable of. It requires hundreds of hours of instruction and thousands of hours of practice to reach a high level of competence. This competence includes not only technical mastery of the instrument of choice, but also a thorough acquaintance with stylistic and expressive norms. That is, musicians must not only know how to play the right notes, in tune, in tempo, with the correct rhythm, and at an appropriate dynamic level, but also how to continuously *vary* tempo and dynamics (and on some instruments, intonation and timbre as well) so as to produce an expression that captures listeners' attention and emotions. This is particularly true of Western art music of the Romantic period, which often calls for extreme modulations in tempo and dynamics which, for the most part, are not notated.

On the piano, expressive timing and dynamics are the two principal dimensions that make a performance interesting and appealing. The present study focuses on expressive timing only. This term denotes continuous modulations of the basic tempo which can be measured and described in terms of the temporal intervals between successive tone onsets (tone interonset intervals or IOIs).[1] The pattern of IOIs, normalized to a fixed nominal note value and plotted as a function of metrical score position, defines the expressive *timing profile* of a performance.[2] A number of earlier studies have analyzed the expressive timing of pianists' performances; for a summary of this literature, see Repp (1992a). In most of these studies, as in the present one, technical difficulty or fingering were not important factors; the music

This research was supported by NIH grant MH-51230. I am grateful to Charles Nichols and Ilan Berman for assistance, to Jonathan Berger (Director, Center for Research in Music Technology, Yale University) for permission to use the Yamaha Disclavier, to José Bowen, Nigel Nettheim, and two anonymous reviewers for helpful comments on an earlier version of the manuscript, and to the pianists for lending their skills to this project.

was relatively slow and easy to play, so that the timing profile was a relatively pure measure of expressive intention (though the precise realization of that intention does require fine motor control!). The most extensive analysis of this kind included 28 recorded performances of Robert Schumann's well-known piano piece, "Träumerei" (Repp, 1992a). Various statistical techniques were applied to assess within-performance consistency, commonalities and differences among individual timing patterns, relationships of timing to musical structure, and the precise shaping of temporal detail (such as *ritardandi*).

The present investigation was modeled closely after this earlier analysis. It was based on a new set of 29 performances by 10 advanced student pianists ("students" henceforth), which had been recorded on a Yamaha Disclavier in MIDI format after only a brief rehearsal. These recordings satisfied the minimal requirements of a professional performance in that they were fluent and expressive. Whereas the "expert" performances studied by Repp (1992a) represent the pinnacle of artistic achievement and insight (in many instances at least), the student performances analyzed here represent expressive performance in a more pristine state, as it were. They result from the relatively spontaneous application of acquired technical and musical skills to a score, albeit one familiar from listening and, in some cases, from past study. Such performances are of interest in their own right because they may be more representative of a standard or norm that guides the expressive shaping of timing (and dynamics) than are the highly individual and refined interpretations of famous concert artists. Whether that is the case, however, is an empirical question that the present study meant to address. In principle, it could well be that student performances are as diverse or more diverse than expert performances, due to large variation in students' competence and degree of musical understanding. Considering the minimal preparation preceding the present recordings, the students' expressive timing might also be more variable than the experts' finely honed profiles, and it might show lapses of control or taste in the shaping of temporal detail. If so, the student performances would probably not be a good basis for studying principles of expressive timing. It will be argued below, on the contrary, that these performances in fact reveal a high level of competence and may actually provide a better starting point for modelling expressive microstructure than the performances of the most famous artists.

The aspects of expressive timing that will be considered may be grouped under four headings: (1) consistency, (2) commonalities, (3) execution of local details, and (4) individual differences.

The pianists' consistency was assessed by comparing their timing profiles across repeated performances and across identical or similar musical passages in the same performance. Consistency may be regarded as a measure of technical precision, provided that the pianist did not intend to play the music differently when it was repeated. A low correlation may indicate a change of interpretation, but since the students had been asked to provide three similar performances, their between-performance consistency was taken as an indication of their ability to reproduce the same expressive intention. A comparison of between- with within-performance consistency was expected to reveal the extent to which the students intended to play repeated sections the same way. High within-performance consistency would also indicate precision of expressive intent and execution.

The group average timing profile gives a picture of what most performances have in common. The relative similarity of the average student and expert profiles was of interest. The students might be expected to show a less differentiated or less varied profile, due to a less thorough structural understanding of the music and a less developed ability to convey that interpretation through expressive timing. As Todd (1985), Palmer (1989), and others have demonstrated, the peaks and valleys in the timing profile are an index of the hierarchical phrase structure of the music, as understood by the performer.

In the execution of local details, the temporal shape of *ritardandi* was of special interest. Several studies have found evidence that the sequence of IOIs during well-executed *ritardandi* tends to follow a parabolic or possibly cubic curve (Kronman & Sundberg, 1987; Repp, 1992a, 1992b; Feldman, Epstein, & Richards, 1992; Epstein, 1995), and Todd's (1995) recent characterization of expressive timing in terms of linear changes in tempo is consistent with that finding also (Todd, submitted). The extent to which *ritardando* timing fits such a curve may thus serve as an index of the individual performer's skill or taste, and the question was whether the students would live up to the examples set by the experts.

Last, but certainly not least, was the issue of individual differences. The experts exhibited various timing strategies, but they were also a very heterogeneous group, representing a wide

range of ages, nationalities, and recording dates. Would the students show similar diversity, or would they be more homogeneous and hark closer to a common norm? This was perhaps the most interesting question of this study, and it was addressed primarily by means of principal components analysis, which gives an indication of the number of statistically independent timing patterns underlying a set of performances.

Although the expert data will be summarized and occasionally reproduced here in a new format, frequent reference will be made to Repp (1992a), particularly its figures and tables. The reader should have a copy of that article available.

METHOD

The music

The score of "Träumerei" is reproduced in Repp (1992a: Figure 1), and a brief analysis is presented there also. The piece consists of six 4-bar phrases, the first two of which are repeated. Phrases 1 and 5 are identical and similar to the (abbreviated) final phrase, whereas phrases 2, 3, and 4 are structurally similar to each other. Positions in the music will be referred to by using the convention "bar-beat -halfbeat "; thus, "15-3-2" refers to the second eighth note of the third beat in bar 15.[3]

The pianists

Nine of the participating pianists were graduate students of piano performance at the Yale School of Music; the tenth was about to enter the same program. Five students were in their first year, one in her second year, and three were third-year students. Their age range was 21 to 29, and they had started to play the piano between the ages of 4 and 8. Seven were female, three male. They will be referred to here by numbers prefaced by the letter P (for pianist).

Procedure

The pianists were sent a copy of the music prior to the recording session. Given their extremely busy schedules, however, most of them came to the recording session without advance preparation. The recording took place in a room housing an upright Yamaha MX100A Disclavier connected to a Macintosh computer which recorded the keyboard and pedal actions in MIDI format. The pianist was given the music and asked to rehearse it at the Yamaha for one hour. There were three other pieces to be played in addition to "Träumerei," about 13 minutes of music altogether. After the rehearsal hour, the pieces were recorded one at a time, in whichever order the pianist preferred, and then the cycle was repeated twice. If something went seriously wrong in a performance, it was repeated immediately. One pianist, P4, as a result of multiple retakes and a computer problem, was able to record only two performances of each piece; all others recorded three, as planned. At the end of the session, each pianist filled out a questionnaire and was paid $50.

The responses to the questionnaire revealed that Schumann's "Träumerei" had been previously studied by three pianists (P5, P7, P8) and played informally by two; the rest knew it well from listening only. The pianists were also asked to indicate how satisfied they were with their performances, choosing from the categories "best effort," "good effort," "average," "below average," and "poor." For "Träumerei," the distribution of choices was 0, 4, 5, 1, 0.

Data analysis

The MIDI data were imported as text files into a Macintosh spreadsheet and graphics program (Deltagraph Professional), where the note onsets were separated from the other events (note offsets and pedal actions) and labeled with reference to a numerical (MIDI pitch) transcription of the score. Only the highest note in each chord received a label, and grace notes were excluded.[4] The labeled note onsets were subsequently extracted, and the IOIs between them were computed. Those IOIs which represented intervals longer than a nominal eighth note were divided into equal eighth-note parts, so that all IOIs represented nominal eighth-note intervals. While this subdivision of longer IOIs is useful for graphic purposes, it may be debated whether they should enter statistical procedures as single or multiple data points. Repp (1992a) used the single data point format in some analyses and also applied a logarithmic transformation to the IOIs. The present analyses instead used the multiple data point format without transformation. Analyses of the expert data were redone in the present format, with minimal differences.

RESULTS AND DISCUSSION

Basic tempo

To begin with, the tempo choices of the students were compared with those of the experts, which provided an opportunity to correct faulty tempo estimates reported in Repp (1992a). Estimating the basic tempo of a performance whose tempo is continuously modulated is not straightforward in view of the asymmetric distribution of IOI durations caused by *ritardandi* at major structural boundaries. Repp (1994a) demonstrated, however,

that listeners' subjective tempo estimates for music *not* containing extreme *ritardandi* (in fact, for the initial 8 bars of "Träumerei") are very close to the reciprocal of the average beat duration (expressed in fractions of a minute). The tempo estimates (beats per minute, or bpm) employed here are therefore based on the average beat duration of bars 1–8 of each performance (including the repeat). The estimates given in Table 3 of Repp (1992a), which were derived by a different method, are almost certainly too high and are superseded by the present estimates. Figure 1 shows these estimated tempi for all expert and student performances, rank-ordered according to average tempo in the case of multiple performances.[5] They span a wide range (from 42 to 67 bpm) but are much slower than the 80 bpm recommended by Clara Schumann in her edition of the music, not to mention the 100 bpm attributed to the composer himself.

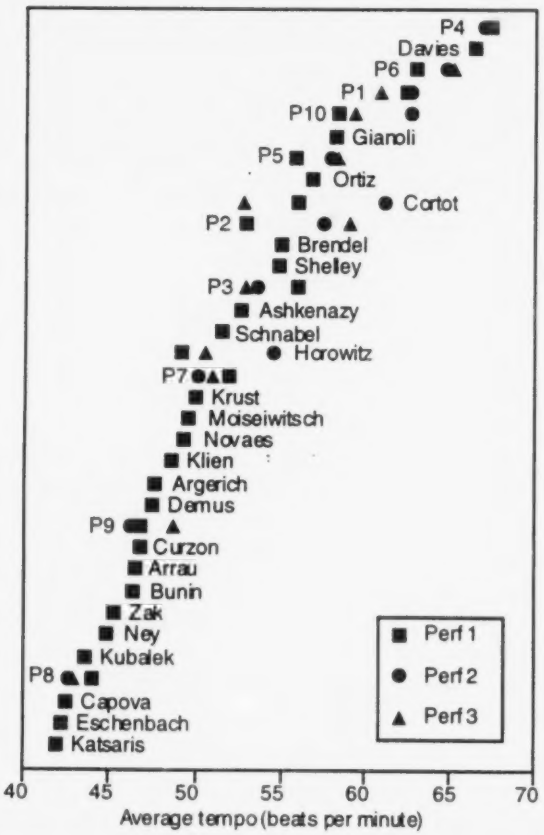


Figure. 1. **Average tempi of the various performances, based on bars 1–8.**

The tempo choices of the student pianists cover as wide a range as those of the experts. Most students, however, turned in relatively fast performances. Those of P4 were slightly faster than the fastest expert performance, which was by Clara Schumann's one-time pupil, Fanny Davies. P6, P1, and P10 played faster than the next-fastest expert, French pianist Reine Gianoli, and P5, P2, and P3 were still in the faster half of the distribution. P7 was near the center, P9 was somewhat on the slow side, and only P8 was near the slow end of the distribution. It can also be seen that no student played all three performances at exactly the same tempo, though P4 (who played only two performances) came close. The two experts who provided three performances each (Cortot, Horowitz) varied more than most student pianists, but their recordings were separated by years whereas the students' performances were only about 20 minutes apart.

One reason for the students' faster tempo choices could have been that they performed the piece in isolation, whereas most of the experts played it in the context of the complete "Kinderszenen" suite. Four of the expert recordings, however, represent performances of "Träumerei" by itself (Katsaris, Capova, Klien, Horowitz-3), and none of them is very fast. Another possible explanation is that the students, especially those who had not studied the piece, were somewhat tense in the recording situation and therefore tended towards faster tempi. The three who had studied "Träumerei" previously (P5, P7, P8) indeed produced some of the slower performances.[6]

Stability of timing patterns across repetitions

High stability of a pianist's expressive timing across repeated performances of the same music seems to be the rule. Although it is often said that artists rarely play the same music the same way twice, or that repeats within a piece should be played differently, such differences seem to be more the exception than the rule with regard to timing. For the student pianists, the stability of timing profiles could be assessed both across repeated performances and across repeats within performances, whereas for the experts (with the exception of Cortot and Horowitz, whose three performances were years apart) only within-performance stability could be assessed.

Within-performance stability can be assessed in three ways in "Träumerei": (a) between bars 1–8 and their repeat ; (b) between bars 1–4 and their literal repeat in bars 17–20; and (c) between bars 9–12 and their almost literal transposition in bars 13–16. The focus will be on the first comparison

here; the others can be made informally in Figure 2 below. Between-performance stability may be determined for complete performances, but in that case the long IOIs associated with major *ritardandi* will have a dominant influence on the correlations; a better choice are bars 1–8, which do not end with an extreme *ritardando*. The comparison of within- and between-performance stability can then also be carried out for bars 1–8.

The between-performance timing profile correlations, averaged over the three pairwise correlations among each pianist's three performances, are shown in Table 1 (columns a and b). Evidently, the student pianists had a high degree of control over their expressive timing patterns. Computed over the whole piece, the correlation was 0.947 on the average. For bars 1–8 alone, the correlations were somewhat lower, due to the absence of very long IOIs, but still quite high (average of 0.907). There is little doubt from these correlations that all students intended to play the piece the same way each time, as they were asked to do. They succeeded in controlling up to 90% of the timing variance, which represents impressive evidence of a cognitive plan that guides rhythmic microstructure.[7]

Table 1. *Average correlations of timing profiles in Schumann's "Träumerei" (a) between entire performances (n = 214), (b) between performances of bars 1–8 only, including the repeat (n = 107), and (c) within performances, between the two renditions of bars 1–8 (n = 53).*

Pianist	(a)	(b)	(c)
P1	0.928	0.871	0.857
P2	0.961	0.924	0.931
P3	0.936	0.916	0.892
P4	0.963	0.925	0.928
P5	0.958	0.913	0.907
P6	0.974	0.917	0.908
P7	0.965	0.903	0.906
P8	0.974	0.953	0.958
P9	0.942	0.877	0.849
P10	0.868	0.867	0.858

The average within-performance correlations for bars 1–8 (Table 1, column c) are as high as the between-performance correlations for the same bars (average of 0.899), indicating that the student pianists did not vary the repeat. The expert pianists' analogous correlations ranged from 0.51 to 0.95 (Repp, 1992a: Table 4, second column).[8] They may have been underestimated slightly, due to human measurement error in the data. Nevertheless, it is clear that the students as a group were as consistent as the most consistent experts. Of course, the experts were free to play the repeat differently, and the lower correlations of some probably reflect such a strategy.[9] Others, however, clearly intended to maintain their original timing pattern; they included such outstanding artists as Arrau, Ashkenazy, and Brendel. P8's correlation of 0.96 may well represent the upper limit of timing accuracy achievable in this portion of the music.

The grand average timing profile

A grand average timing profile was obtained by first averaging the timing profiles (sequences of IOIs) of each student pianist's three performances (two in the case of P4) and then averaging across all 10 students. Subsequently, the profiles for the two renditions of bars 1–8 were averaged, as they were extremely similar and did not differ in average tempo. The resulting timing patterns are shown in Figure 2, which uses the format established in Repp (1992a: Figure 3). Eighth-note IOI durations are plotted on a logarithmic scale to reduce the graphic excursions of the large *ritardandi* and make the detailed variation of the shorter IOIs more visible. The abscissa shows bar and beat numbers for bars 1–4 and 5–8 (left and right panels), corresponding to the initial two 4-bar phrases. The patterns for the remaining four phrases are overlaid on those of bars 1–4 and 5–8; the bar numbers of the abscissa need to be incremented accordingly. IOIs longer than a nominal eighth note are represented as multiple data points or "plateaus."

The left panel of the figure shows that the average timing profiles of bars 1–4 and 17–20 virtually coincide. Since they represent identical music, this is another illustration of the student pianists' high consistency. Bars 21–24, which start out similarly, soon deviate because of the approach to the *fermata* in bar 22; then, from the middle of bar 23 onwards, the final *ritardando* holds sway. In the right panel of the figure, the close timing similarity of bars 9–12 and 13–16 may be observed; these phrases are notationally almost identical but in different keys. Their profiles deviate only during the second half of the last bar, where bar 16 exhibits a greater *ritardando* than bar 12, due to the "deeper" structural boundary following bar 16. The profile for bars 5–8 partially coincides with those of bars 9–12 and 13–16, precisely where the musical material is highly similar. It deviates at points of structural difference, particularly in the second and fourth bars of the phrase, where its local tempo is slower. The final *ritardando* is highly similar in bars 8 and 12.

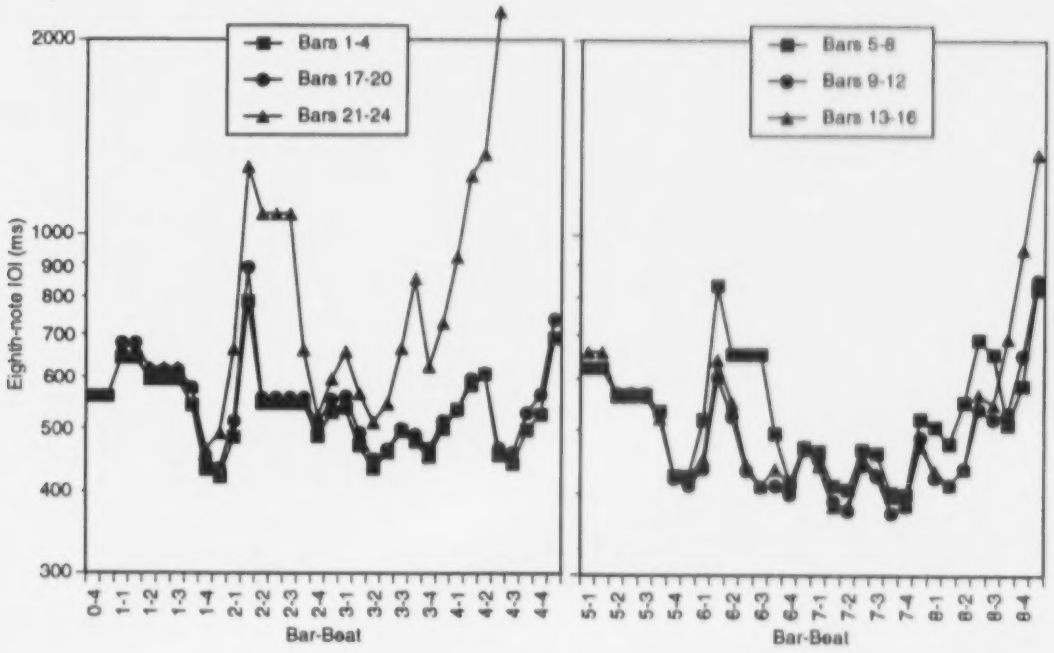


Figure. 2. **Grand average timing profile of the 10 student pianists. Structurally similar phrases are superimposed.**

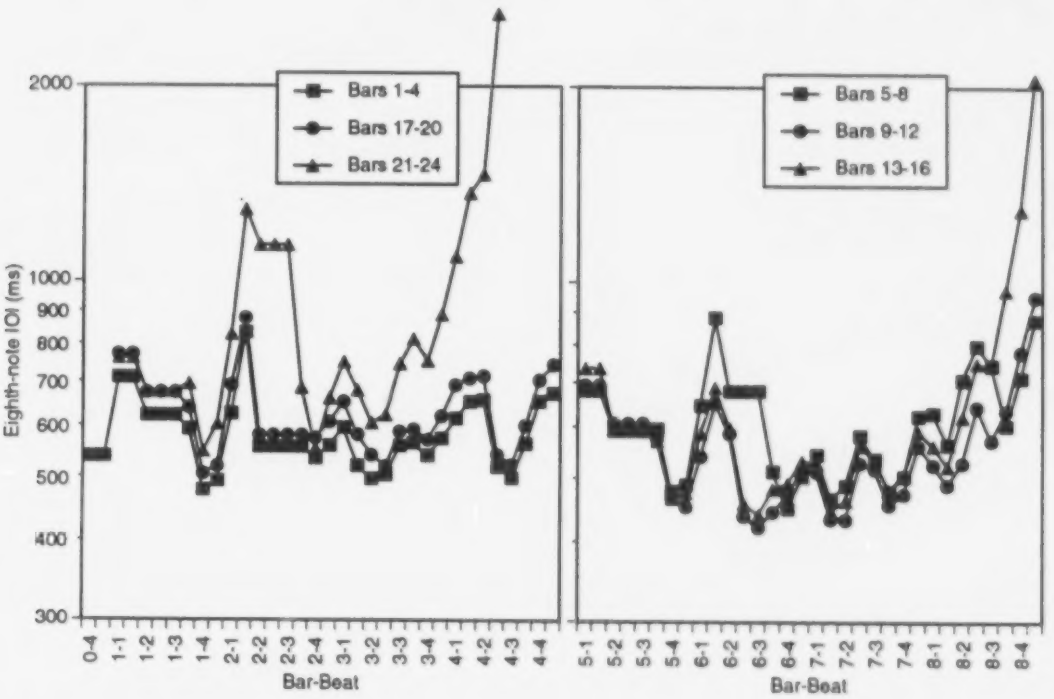


Figure. 3. **Grand average timing profile of the expert pianists.**

The students' average timing profile may be compared with that of the experts, which is reproduced in Figure 3.[10] The similarity is quite remarkable; in fact, there is not a single qualitative difference between the two profiles. One must look carefully to detect some small quantitative differences: Apart from playing somewhat slower overall than the students, the experts tended to play bars 17–20 a little slower than bars 1–4 whereas the students did not; their initial upbeat was a little shorter relative to the following chords than the students'; they lengthened the IOIs less in positions 4-4-2, 20-4-2, and 23-3-2; and they slowed down a little earlier and made a more pronounced *ritardando* in bar 16 compared to bar 12. The correlation between the student and expert grand average profiles was 0.964, and that between their profiles for bars 1–8 only was 0.922. These correlations show that, *on the average,* the students played with almost exactly the same expressive timing as the experts. This is especially remarkable in view of the fact that most of the students had practiced the piece only for a few minutes.

Intercorrelations among performances

The timing similarities among all individual pianists' performances were assessed by computing the correlations between their timing profiles. As already mentioned, when such correlations are computed over all IOIs, the very long IOIs associated with major *ritardandi* dominate and lead to high correlations of restricted range, since all pianists mark major phrase boundaries in this manner. It was more informative, therefore, to examine the intercorrelations for bars 1–8 only (including the initial upbeat), where very long IOIs are absent.

Inspection of the intercorrelation matrix for experts and students combined (n = 38) revealed that all students showed high correlations with all other pianists' performances, except with Argerich, Bunin, Horowitz, Moiseiwitsch, and especially Cortot. These highly individual artists in turn showed lower correlations with other experts' performances. For each pianist's performance, the three most highly correlated performances were determined.[11] This tally revealed that the students' performances were more similar to each other than to the experts' performances. Twenty-two of the 30 correlations (73%) represented other students, even though they constituted only 9/37 = 24% of the possible candidates. Of the 8 expert performances in this set, four were by one pianist (Capova) and two by another (Ashkenazy). In fact, for five students the most similar expert performance was that of Capova, and for three that of Ashkenazy. All ten students showed relatively high correlations with Capova, eight with Ashkenazy, and nine with Zak, another Russian pianist.[12]

Conversely, and perhaps more surprisingly, the experts' timing profiles tended to correlate more highly with the students' profiles than with those of other recording artists. Fifteen of the 28 expert performances correlated most highly with a student performance. Forty-two (50%) of the 84 "highest three" correlations were with students, even though they constituted only 10/37 = 27% of the candidates. Nine student pianists were represented among those correlations; the one absent was P8, who had conspicuously lower correlations with most experts' performances. Two factors may account for the experts' higher correlations with students than with other experts: First, it is possible that the students' profiles were more representative of the grand average timing profile, so that they were closer, on the average, to most other performances than were the more eccentric expert performances. This will be investigated further below. Second, the absence of measurement error and the reduction of random variation by averaging over three performances made the student profiles statistically more reliable than the expert profiles, which contained some random error; this may have enhanced the correlations with the student performances. It should also be noted that the expert performances' correlations with other performances were generally lower than the students' correlations with other performances.

Finally, it was evident that each student's own three performances were more similar to each other (Table 1) than their average was to any other pianist's performance. Likewise, as already noted by Repp (1992a), Cortot's and Horowitz's respective three performances, even though they had been recorded years apart, were more similar to each other than to any other pianist's performance. Thus, each pianist, whether expert or student, seems to have a replicable "timing signature," part of his or her individuality. However, the similarity to other pianists' performances may be nearly as great.

Extent of timing variation

Correlations are sensitive only to differences in shape among the timing profiles, not to the

magnitude of the expressive tempo modulation (a scale factor). A measure of the range of individual timing variation is the standard deviation of the IOIs, or even better, the coefficient of variation, which is the standard deviation divided by the mean. This measure corrects for the natural tendency of absolute timing variation to increase at a slower tempo (cf. Repp, 1994b); thus it is a measure of relative timing variation. Individual coefficients of variation for bars 1–8 ranged from 0.14 to 0.31 across the 38 performances. Both the least modulated (Schnabel, Klien, Eschenbach) and the most highly modulated performances (Argerich, Demus, Horowitz-3, Cortot-1, Arrau) were by experts; the students' values extended over a narrower range from 0.17 (P7) to 0.24 (P3).

Principal components analysis

Following Repp's (1992a) procedures, principal components analysis (PCA) with Varimax rotation was employed to determine whether more than one shared pattern of variation underlies the individual pianists' timing profiles.[13] The analyses were conducted on expert and student profiles combined as well as separately.[14] Analyses conducted on the complete performances yielded single-component solutions, which basically showed that all pianists marked the major phrase boundaries with *ritardandi*, even though there were individual differences in their extent. It was again more informative to restrict the analyses to bars 1–8 (including the initial upbeat), which did not contain any extreme *ritardandi* and thus permitted the performance intercorrelations to vary over a wider range.

The PCA on the experts yielded results similar to those reported in Repp (1992a: Table 5, Figure 4), except that the fourth component fell just short of significance. Three significant components (i.e., with eigenvalues greater than 1) accounted for 72% of the variance. Before rotation, the first principal component (a kind of grand mean) accounted for most of this variance, but Varimax rotation redistributed this variance among the components so as to maximize discrepancies among component loadings, which usually facilitates interpretation. The first rotated component still represented the most common timing pattern, with Schnabel showing the highest loading (profile-component correlation) by far and many other pianists showing substantial loadings. The second component was the "Horowitz component," with several other pianists (especially Argerich) showing moderate loadings, and the third component was virtually unique to Cortot.

The separate analysis on the students, by contrast, yielded only a single significant component that accounted for 80% of the variance. Component loadings ranged from 0.94 (P2) to 0.82 (P8). Thus it is evident that the students showed much less individual variability than the experts.

In the combined analysis of experts and students, five significant components emerged which together accounted for 80% of the variance. Before Varimax rotation, the first principal component explained 62%, and the others 7%, 5%, 3%, and 3%, respectively. Clearly, the first component was again a sort of grand mean, and all individual performances loaded on it with values of 0.51 or higher. The student pianists in particular had high loadings (0.73–0.91), and among the experts Capova (0.93) and Ashkenazy (0.86) had the highest loadings. This confirms that these pianists' timing profiles were all close to the grand average timing pattern. The five rotated components accounted for 25%, 20%, 14%, 11%, and 10% of the variance, respectively. The rotated component loadings of the 38 performances are shown in Table 2. Values of less than 0.4 are omitted.

The first of these rotated components was similar to the grand average, but there was a much wider range of component loadings now. Still, 8 of the 10 student pianists had their highest loading on this component, and even the remaining two (P3, P10) showed a modest correlation with it. Interestingly, P8, who seemed least typical in the separate analysis on the students, had the highest loading of all. Expert pianists represented most strongly by the first component were Kubalek, Schnabel, Zak, Capova, and Davies. The second component was the "Horowitz component." Other moderately high loadings were all by experts; only three students showed small correlations (between 0.4 and 0.5) with this component. The third component was new (compared to the analysis on the experts alone), and six experts as well as two students had their highest loadings on it, with three additional students showing small correlations. The highest expert loadings were by Katsaris, Demus, Ashkenazy, and Shelley. The fourth component was the idiosyncratic "Cortot component," with no students and only two other experts minimally represented. The fifth component, defined by Novaes and Brendel, among others, also showed little student representation.

Table 2. *Component loadings of the Varimax-rotated 5-component solution for bars 1–8. Only correlations above 0.4 are listed; the highest loading of each performance is in bold face.*

Pianist	Comp I	Comp II	Comp III	Comp IV	Comp V
P8	**0.831**				
Kubalek	**0.751**				
P7	**0.750**				
Schnabel	**0.710**				0.436
P2	**0.690**				0.462
Zak	0.677	0.442			
P9	**0.672**	0.431			
Capova	**0.665**	0.409			
P5	**0.664**		0.499		
Davies	**0.656**				
P6	**0.630**				0.406
P4	**0.603**		-0.522		
P1	**0.572**	0.464	0.411		
Ortiz	**0.532**	0.400			
Horowitz-2		**0.858**			
Horowitz-3		**0.850**			
Horowitz-1		**0.834**			
Argerich		**0.727**			
Eschenbach		**0.658**			
Klien	0.514	**0.6[illegible]**			
Gianoli	0.413	**0.575**			
Ney		**0.563**			0.551
Katsaris			**0.697**		
P3	0.423		**0.670**		
Demus		0.434	**0.669**		
P10	0.445	0.495	**0.575**		
Ashkenazy	0.455		**0.559**		
Shelley			**0.557**		0.496
Bunin		0.496	**0.529**		
Cortot-3				**0.873**	
Cortot-1				**0.873**	
Cortot-2				**0.868**	
Novaes		0.509			**0.619**
Brendel	0.463	0.457			**0.564**
Moiseiwitsch	0.491				**0.539**
Krust	0.444			0.481	**0.536**
Arrau	0.421	0.406	**0.474**		
Curzon	**0.432**			0.422	

These results confirm the general impression that the students were relatively conservative in their timing patterns and stayed close to the most representative timing profile, though the combined analysis revealed some influence from Katsaris and Horowitz type patterns (Components III and II, respectively). Several student pianists actually mentioned that they had been impressed and possibly influenced by Horowitz's famous performances of "Träumerei" (his favorite encore). On the whole, however, the students were decidedly "mainstream" in their timing strategies.

The components just discussed represent abstractions from the data. Each rotated component represents a particular "underlying" timing profile that is only similar to, but not identical with, certain individual performances. Each individual performance can be represented as a linear combination of these underlying (orthogonal) profiles, plus variation unique to the performer (i.e., variance not accounted for). Rather than focusing on these abstract patterns (see Repp, 1992a: Figure 4), we will now examine individual differences in the execution of detailed timing maneuvers.

Upbeats

Each phrase in "Träumerei" begins with an upbeat, which appears in the score as an unaccompanied quarter note (bar 0), an accompanied quarter note (bars 4 and 20), an accompanied eighth note (bars 8 and 12), or a grace note (bar 16). It is followed by a melody note an interval of a fourth higher (accompanied by a bass note) and, on the next beat, by a four-note

chord, which is in turn followed after one beat and a half by a melody note that initiates the next melodic gesture. In Repp (1992a: Figure 5), the expert pianists' timing behavior was portrayed in terms of two ratios among the three normalized IOIs (A, B, C) defined by these four events.[15] The first ratio, 5A/(2B+3C), describes the timing of the upbeat relative to the 5-eighth-note IOI between the two subsequent melody notes, whereas the second ratio, B/C, describes the relative placement of the chord in this long IOI. In each case, a ratio of 1 implies that the local tempo (the underlying beat) remained constant. The students' ratios are shown in Table 3.

The expert pianists had a strong tendency to shorten the initial unaccompanied upbeat (bars 0–1) by various degrees; their ratios ranged from about 0.4 to 1.1. The student pianists showed a similar tendency, but their range of ratios was much more restricted (0.83 to 1.05) and thus closer to the nominal value of the quarter-note upbeat. The experts' relative timing of the accompanied quarter-note upbeats in bars 4–5 and 20–21 also showed a wide range (ratios from about 0.7 to 1.4), with a slight tendency towards lengthening the upbeat, due to its overlap with the end of the preceding phrase. The students were again more conservative, with ratios between 0.90 and 1.22. The eighth-note upbeats in bars 8–9 and 12–13 were almost always lengthened by the experts, again due to their straddling a phrase boundary; the ratios ranged from 0.7 to 1.8 in bars 8–9, and from 0.8 to 2.4 in bars 12–13. The students also exhibited lengthening and considerable variation here, with ratios from 1.09 to 1.99 in bars 8–9, and from 1.09 to 2.20 in bars 12–13. The most interesting upbeat is the grace note in bar 16 (considered nominally an eighth note here), which in expert interpretations ranged from the equivalent of a sixteenth note (0.4) to that of a quarter note (2.2), with many gradations in between. The students, too, showed a variety of ratios ranging from 0.52 to 1.66. P1 and P5 played the grace note effectively as a sixteenth note, P6 and P8 as a literal eighth note, and the others as an eighth note lengthened by various degrees.

As to the relative placement of the following four-note chord, expert pianists showed an overwhelming tendency to lengthen the preceding (two-eighth-note) IOI relative to the following(three-eighth-note) IOI in all positions in the music, though there were some exceptions, most notably Cortot (Repp, 1992a: Figure 5). Their B/C ratios were generally between 0.8 and 1.6. Here the students show again a more restricted range, from 0.95 to 1.34 in terms of average ratios (from 0.88 to 1.49 in terms of ratios for individual instances). It should be acknowledged that previous averaging over the students' three performances may have contributed to a relative narrowing of their range of ratios in these comparisons. However, the students never exhibited the rather eccentric upbeat timing of some of the famous pianists (such as Cortot or Argerich).

The ascent to the melodic peak

The apex of each phrase is reached by an ascending melodic gesture composed of 5 eighth notes and a long note, thus comprising 5 IOIs. The gesture recurs eight times in the music, or six times after averaging over the two renditions of bars 1–8. Repp (1992a) showed that, for most expert pianists, the sequence of IOI durations in this gesture can be described by a quadratic (parabolic) function that first descends and then ascends to the pitch peak. As Todd (1992, 1995) has shown, this time course corresponds to physical motion with a constant deceleration and acceleration.

Table 3. *Ratios 5A/(2B+3C) and B/C of the three normalized IOIs between the first four events in each phrase. The B/C ratios are averaged over all 8 occurrences, the first two 5A/(2B+3C) ratios over the two renditions of bars 1–8. The grace note in bar 16 is considered a nominal eighth note.*

	5A/(2B+3C)						B/C
Bars:	0–1	4–5	8–9	12–13	16–17	20–21	
P1	0.92	1.17	1.49	1.18	0.63	1.10	1.34
P2	0.93	1.06	1.42	1.73	1.31	0.98	1.07
P3	1.05	1.14	1.99	2.20	1.62	1.22	1.22
P4	0.89	1.14	1.55	1.52	1.66	1.07	1.29
P5	0.83	0.90	1.15	1.09	0.52	0.89	1.04
P6	0.91	1.03	1.47	1.31	0.92	1.02	1.19
P7	0.87	0.98	1.09	1.19	1.14	0.91	1.01
P8	0.83	0.92	1.27	1.16	0.98	0.97	0.95
P9	0.98	0.99	1.34	1.42	1.18	0.99	0.96
P10	0.95	1.14	1.47	1.59	1.30	1.14	1.31

Thus, pianists' ability to follow such a timing curve may be taken as an index of their ability to shape a phrase, as long as it can be assumed that no atypical timing pattern was intended. In the case of Cortot, who consistently shortened the last IOI (except in bars 21–22), this assumption was clearly not warranted, and a few other pianists (Argerich, Bunin, Curzon) occasionally showed a pattern similar to Cortot's. The large majority of experts, however, showed good to excellent fits to a parabolic curve, with individual variations in its elevation and degree of curvature.[16]

The students' average timing functions for the six instances of the melodic gesture are shown in Figure 4. They are quite similar in pattern to the averages of the expert pianists (Repp, 1992a: Figure 6), and the quadratic fits are satisfactory (r^2 ranged from 0.970 to 0.985, versus 0.949 to 0.999 for the experts), though it may be noted that the fourth IOI tends to be too short. This was mainly due to one pianist's atypical functions, just as a tendency for the fifth IOI to be too short in the average expert data was mainly due to Cortot and a few others.

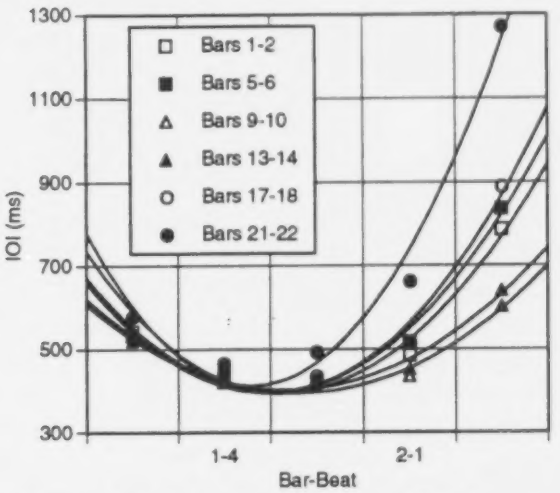


Figure 4. **Parabolic curve fits to the students' average timing patterns in the "ascent to the melodic peak" in six different positions in the music.**

The student pianists' individual fits to parabolic timing curves were good to excellent. Only P4 repeatedly showed a tendency to shorten the fourth IOI, resulting in fits (r^2 values) between 0.851 and 0.922. P7 also showed relatively poor fits in the middle section (r^2 = 0.916, 0.804). All other goodness-of-fit indices ranged from 0.924 to 0.999. By comparison, the experts' performances included some much poorer fits, ranging from 0.027 to 0.694 for Cortot (except in bars 21–22), for example, and from 0.073 to 0.779 for Curzon in the same positions. These distinguished pianists apparently had some unconventional ideas about how the melodic gesture should be shaped. However, most of the experts' fits were in the same range as the students', suggesting that the students (with the possible exception of P4) had mastered the art of shaping a phrase in the conventional manner.

The ranges of curvatures of the parabolic functions may also be compared. Repp (1992a: Figure 9) plotted average curvatures in bars 1–2, 5–6, and 12–18 against those in bars 21–22 for those cases that yielded a good parabolic fit; the respective ranges (the values represent the coefficients of quadratic equations, with the five IOIs numbered serially) were 30 to 160 and 20 to 150, respectively. The comparable ranges for the students were 33 to 90 and 70 to 168, respectively. These ranges were more restricted, but this difference was mainly due to Demus's extremely high curvatures, and to Horowitz's abnormally flat functions in bars 21–22. For the most part, experts and students showed similar degrees of curvature, or temporal inflection.

The average relative timing of the long IOI at the melodic peak, even though it varied with context, was extremely similar for experts and students, as can be seen in Figures 2 and 3.

Grace notes

The final long note of the melodic gesture just discussed (the peak of the phrase) is preceded by two grace notes in the left hand, basically a written-out *arpeggio*, in bars 2, 6, and 18. The conventional way of playing this passage, suggested by the notation, is to fit the two grace notes into the preceding IOI, so that the third and final left-hand note coincides with the melody note (and its lower octave) in the right hand. Thus the IOI can be divided into three portions, A, B, and C, and ratios can be calculated which reflect the relative placement of the first grace note in the IOI, A/(B+C), and the relative "durations" of the two grace notes, B/C.[17] Repp (1992a: Figure 10) discovered that only about half of the experts played this passage the conventional way. At least five variants were observed, some of which made it impossible to calculate the ratios just described. For those instances where the ratios could be determined, they varied widely but tended to cluster around modal values of 0.4 and 0.5 respectively (Repp, 1992a: Figure 11). In other words, the first grace note tended to occur after

about 30% of the IOI had elapsed, and the second grace note shortly after the middle of the IOI.

The students, too, did not all play the passage the conventional way. P7 played the first grace note at the same time as the preceding melody note (as did Schnabel), so that the A/(B+C) ratio was close to zero. P10, on the other hand, played the second grace note very close to the following melody note (as did Bunin and Kubalek), so that the B/C ratio became very large. Both P8 and P10 delayed the top note in the left hand and P9 advanced it, but this did not affect the ratios, since the IOI was measured to the onset of the melody note in the right hand. With P7 and P10 omitted, the student pianists' average ratios ranged from 0.22 to 0.57 for A/(B+C), and from 0.33 to 0.77 for B/C.[18] These ranges coincide precisely with the main cluster of expert values (Repp, 1992a: Figure 11). Again, the students can be seen to be similar to the experts in their microtiming but more conservative on the whole, as a larger number of deviant ratios and strategies was observed among the experts.

The music contains two other, isolated grace notes. One of these, the upbeat in bar 16, has already been discussed. The other one is the melodic grace note in bar 8, whose onset falls within the fourth eighth-note IOI (position 8-2-2) in that measure. Most expert pianists played it near the middle of the IOI; its relative position ranged from 33% to 70%. Two pianists (Cortot-3, Davies) omitted the grace note, and in other performances Cortot played E instead of C. (One must be a very great pianist to take that kind of liberty!) Two student pianists also were deviant: P10 consistently omitted the grace note, whereas P9 sustained the grace note and thus omitted (or effectively advanced) the following melody note. For the other 8 student pianists, the average relative timing of the grace note ranged from 33% to 42% (from 27% to 46% across individual instances). Here, at last, is a clear difference between experts and students: The students tended to play the grace note earlier. Only four expert pianists (Capova, Cortot-2, Gianoli, Krust) fell within the students' range of timing.

The descent from the melodic peak

The second half of each phrase consists of a series of falling melodic gestures or "phraselets." There are two versions of this descent, one instantiated in bars 2–4, 18–20, and 22–24, and the other in bars 6–8, 10–12, and 14–16. They will be referred to as Type A and Type B, respectively. The average timing patterns for the three occurrences of each type were very similar, except for varying extents of the phrase-final *ritardandi* (see Figs. 2 and 3). To obtain a more detailed picture of individual variability in the temporal shaping of these complex passages, they were subjected to separate principal component analyses.

The analysis of the Type A phrases did not include bars 22–24 because the very long IOIs of the final *ritardando* would have dominated the intercorrelation structure. The relevant IOIs thus ranged from position 2-4-1 to 4-4-2, and from 18-4-1 to 20-4-2. The PCA on the experts yielded six significant components which accounted for 82% of the variance. The analysis on the students, however, yielded only a single component, accounting for 76% of the variance. This provides a particularly striking demonstration of the greater homogeneity of the students' timing patterns. The combined analysis yielded seven components, accounting for 86% of the variance. After Varimax rotation, the variance was distributed among the components as follows: 23%, 15%, 15%, 10%, 9%, 9%, and 4%. The component loadings are shown in Table 4. Six of the components resembled those obtained for the experts alone (cf. Repp, 1992a: Table 6), though their order had changed somewhat. The new component, remarkably, was the first and most important one. All 10 student pianists, but only three experts (Capova, Bunin, Shelley), had their highest loading on this component. This suggests that the students as a group represented a particular style of timing in this musical passage. Two students showed modest loadings on the second component (the "Horowitz component"), six showed affinities with the third component (defined by Ney and others), and one with the fourth component (the "Cortot component"). The remaining three components, though they showed high loadings by some expert pianists (Moiseiwitsch; Novaes and Schnabel; Argerich) had no significant student representation.

The timing profiles corresponding to Components II–VI, reconstructed from the standardized component scores, can be seen in Repp (1992a: Figure 12).[19] The new Component I pattern is shown in Figure 5. This underlying timing profile is characterized by relatively even timing through bar 3, with only slight *ritardandi* during the first two phraselets, followed by a huge *ritardando* towards the end of the third phraselet (positions 4-1-2 and 4-2-1), and a pronounced lengthening of the last IOI, which marks the end of the fourth phraselet in the bass voice and accompanies the upbeat to the following phrase.[20]

Table 4. *Component loadings of the Varimax-rotated 7-component solution for bars 2-4-1 to 4-4-2 and 18-4-1 to 20-4-2. Only correlations above 0.4 are listed; the highest loading of each performance is in bold face.*

Pianist	I	II	III	IV	V	VI	VII
P7	**0.874**						
Capova	**0.823**						
P8	**0.798**						
P6	**0.759**		0.464				
P5	**0.738**						
Bunin	**0.683**						
P2	**0.681**		0.533				
P9	**0.646**	0.535					
P3	**0.645**		0.622				
P5	**0.634**		0.504	0.420			
Shelley	**0.594**		0.426		0.563		
P1	**0.589**		0.435				
P10	**0.507**	0.427	0.443				
Horowitz-2		**0.912**					
Horowitz-1		**0.859**					
Horowitz-3		**0.783**					
Klien		**0.767**					
Zak	0.503	**0.625**					
Demus		**0.592**					
Ney			**0.775**				
Eschenbach		0.443	**0.684**				
Ashkenazy			**0.681**				
Brendel			**0.651**				
Katsaris	0.472		**0.638**				
Arrau	0.551		**0.626**				
Cortot-1				**0.851**			
Cortot-2				**0.810**			
Cortot-3				**0.683**			0.462
Ortiz		0.471		**-0.659**			
Moiseiwitsch					**0.876**		
Davies					**0.578**		
Gianoli	0.462			0.417	**0.559**		
Krust					**0.510**	0.428	
Novaes						**0.837**	
Schnabel						**0.829**	
Kubalek	0.560					**0.686**	
Curzon					0.554	**0.596**	
Argerich	0.463	0.445					**0.608**

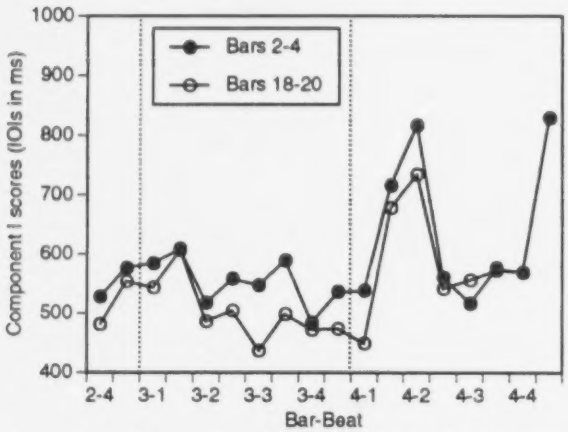


Figure 5. **Timing pattern of Component I for the "descent from the melodic peak" in Type A phrases in the combined analysis of students and experts.**

All students exhibited high within-performance consistency by playing bars 2-4 and bars 18-20 very similarly. Most experts, on the other hand, played bars 18-20 more slowly than bars 2-4, and their timing profiles for these two musically identical passages also often diverged considerably, especially in Horowitz's and Cortot's renditions. There were some experts, however, who, like the students, played the two passages almost identically (most notably Arrau, Capova, Kubalek, and—surprisingly—Horowitz-3).

The PCA of the Type B phrases included positions 6-4-1 to 8-1-2, 10-4-1 to 12-1-2, and 14-4-1 to 16-1-2. The final six IOIs of each phrase had to be excluded because of the large *ritardandi* they carried, which will be analyzed separately below. The analysis of the expert performances again yielded six components, accounting for 80% of the vari-

ance, whereas the student analysis once again gave rise to only a single significant component, accounting for 65% of the variance. Eight components emerged in the combined analysis, accounting for 84% of the variance. The component loadings for the first 7 components are shown in Table 5. Again, the addition of the students to the experts resulted in a new first component. The five highest loadings on that component represent student pianists; four additional students had loadings above 0.4. No expert loaded very highly on this component; Arrau, Capova, and Argerich showed moderate correlations with it. The other novel component was Component V, defined by Curzon and Schnabel. Two students had small loadings on this component, though for one of them (P1) it was the highest on any component.[21] The remaining six components resembled those obtained in the analysis on the experts alone, with the order of the first two components reversed (see also Repp, 1992a: Table 7). Two students (P3, P6) had their highest loadings on Component II, defined by Krust and two of Cortot's performances; P2 loaded weakly. No student was represented on Component III, associated most strongly with Klien and Moiseiwitsch, as well as with two of Horowitz's performances. Four students were represented on Component IV, defined by Zak, Ney, and Davies. Only P1 was marginally associated with Component VI, which was almost unique to Ortiz. Component VII, defined by Bunin and Shelley, and Component VIII, unique to Cortot-3 (with a loading of 0.818), showed no student affiliation.

Table 5. *Component loadings of the Varimax-rotated 8-component solution for bars 6-4-1 to 8-1-2, 10-4-1 to 12-1-2, and 14-4-1 to 16-1-2. Only correlations above 0.4 are listed; the highest loading of each performance is in bold face. Component VIII is not shown.*

Pianist	I	II	III	IV	V	VI	VII
P4	**0.833**						
P8	**0.823**						
P9	**0.777**						
P2	**0.765**	0.434					
P10	**0.587**			0.511			
Capova	**0.585**					0.438	0.408
Argerich	**0.527**		0.470				
Krust		**0.827**					
Cortot-1		**0.775**					
Cortot-2		**0.742**		0.460			
P3	0.437	**0.630**			0.401		
Brendel		**0.593**			0.407		
Kubalek		**0.586**					
Novaes	0.455	**0.568**					
P6	0.509	**0.530**		0.521			
Klien			**0.777**				
Moiseiwitsch			**0.732**				
Arrau	0.629		**0.678**				
Horowitz-2			**0.677**				
Horowitz-3	0.435		**0.665**	0.400			
Gianoli		0.427	**0.568**				
Zak				**0.727**			
Ney				**0.716**	0.469		
Davies				**0.656**			
Curzon		0.411			**0.732**		
Schnabel					**0.726**		
Katsaris				0.574	**0.598**		
Ortiz						**0.889**	
Eschenbach	0.412					**0.509**	
Bunin							**0.691**
Shelley		0.411			0.537		**0.548**
Cortot-3							
Horowitz-1				0.407			0.426
P5	0.422			0.456			
Demus			0.403	0.457			
Ashkenazy			0.490				
P1					0.495	0.408	
P7	0.407			0.425			

The diversity of patterns in this part of the music is noteworthy, with even Horowitz's and Cortot's performances varying significantly amongst themselves. Most pianists were not strongly associated with any single component but showed contributions from several; this was true for the students as well as the experts. Nevertheless, the students did tend to cluster on Component I. It should also be noted that the similarity structure captured by the component loadings is entirely different from that for the Type A phrases; there are hardly any expert pianists that "stayed together" in terms of their primary component affiliations (exceptions are Klien and Horowitz; and Capova, who stayed with the students). The distribution of the variance accounted for among the rotated components was also less skewed than in the Type A analysis, which indicates that there were no strongly dominant patterns.

The timing pattern associated with Component I is shown in Figure 6. It shows a clear (and representative) tendency for the phrases in the middle section (bars 10–12, 14–16) to be played faster than bars 6–8. Otherwise, the timing profiles are fairly parallel, showing a tendency to accelerate during the phrase (before the final *ritardando*) and a pronounced lengthening of the IOIs preceding downbeats (positions 6-4-2 and 7-4-2), which also precede salient harmonic changes.

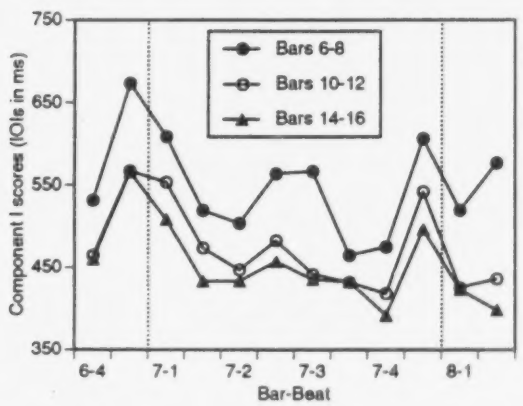


Figure 6. **Timing pattern of Component I for the "descent from the melodic peak" in Type B phrases in the combined component analysis of students and experts.**

Again, the students showed considerable similarity among each other, while the expert profiles were much more diverse. Most students played the three instances of the Type B phrase similarly, whereas experts more often varied their timing. The secondary components which influenced the students most, II and IV, differed from Component I more in overall trend than in qualitative detail, hence the apparent homogeneity of the student group. Component III, which showed a pronounced lengthening of downbeat rather than pre-downbeat IOIs, was peculiar to a small group of experts (including two of Horowitz's performances), as were Components V and VI.

K. Phrase-final ritardandi

There are three major ritardandi in the piece, in bars 12, 16, and 23–24, respectively. The ones in bars 12 and 16 comprise four IOIs each, whereas the final ritardando exhibits a structural and agogic break (a "comma" in the score) that sets the final two IOIs apart from the four preceding ones (cf. Figures. 2 and 3). Repp (1992a: Figure 14) found that each of the 4-IOI progressions was fit very well by a parabolic function, at least when the average across all expert pianists was considered. Individual fits were not always so close, especially in bar 12, but satisfactory on the whole. These fits were re-examined quantitatively here, to compare them to those for the students.

Figure 7 shows the parabolic fits to the students' average ritardando curves. All three fits were remarkably close, with r^2 greater than 0.999 in each case. The curvature of the functions was greater than that of the experts' average functions in bars 12 and 23 - 24, but less in bar 16.

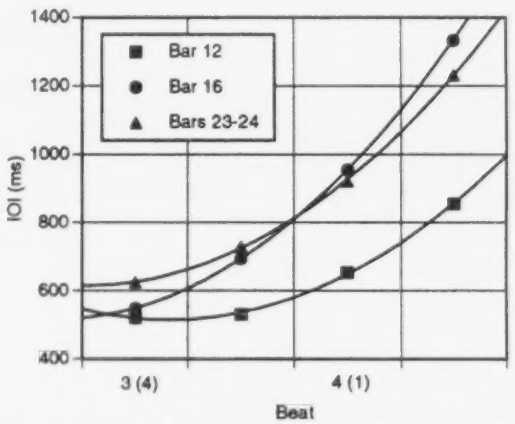


Figure 7. **Parabolic curve fits to the students' average timing patterns in the major phrase-final *ritardandi*. The beat numbers in parentheses refer to bars 23–24.**

To compare the individual fits, the following classification of r^2 values was made: excellent (> 0.999), very good (0.99 - 0.999), good (0.95 - 0.99), moderate (0.90 - 0.95), poor (0.80 - 0.90), and unacceptable (< 0.80). Table 6 shows for each of the three *ritardandi* the distribution of fits among these categories for experts and students.

Table 6. *Comparison of r^2 values of parabolic fits to ritardando functions.*

Quality of fit	Bar 12		Bar 16		Bars 23 - 24	
	Experts	Students	Experts	Students	Experts	Students
$r^2 > 0.999$	6	0	5	2	3	2
0.99 - 0.999	4	4	11	5	10	5
0.95 - 0.99	6	2	10	1	8	3
0.90 - 0.95	4	1	2	2	4	0
0.80 - 0.90	5	3	0	0	1	0
$r^2 < 0.80$	3	0	0	0	1	0
Min. curvature	-98	-15	7	-80	-27	-21
Max. curvature	117	76	447	183	239	236

They are quite comparable. Both groups showed poorer fits in bar 12 than in the other locations. Instances of unacceptable fits were observed only among the experts, as were excellent fits in bar 12; thus the experts again exhibited somewhat greater diversity.

The range of curvatures (considering only fits of better than 0.90) was also greater for the experts than for the students in bars 12 and 16. Among students and experts alike, there were instances of negative curvature, i.e., *ritardandi* about to change into *accelerandi*. However, no pianist exhibited such a function in all three positions.

Finally, the last two IOIs (positions 24-2-1 and 24-2-2) were examined in terms of their ratio. Among the experts, the ratios varied from 1.22 to 2.24, among the students from 1.33 to 1.93—again a somewhat smaller range.

GENERAL DISCUSSION

The performances of Schumann's "Träumerei" examined here were obtained under conditions that—some might argue—make a comparison with commercially recorded expert performances futile. Not only were the student pianists younger, less experienced, and probably less talented than the famous pianists, but they also were less prepared, played from the score on a mediocre piano, and even committed some errors (though none that affected timing). It is the more remarkable, therefore, that the students were fully the equal of the experts in terms of measures of timing precision and consistency. If anything, they were *more* consistent than the experts, since the least consistent artists were all from the expert camp. What this demonstrates is that even a minimally prepared performance by a competent pianist has a precisely defined underlying plan that governs its expressive timing pattern. This plan presumably derives from tacit knowledge of general rules of expressive timing that can be implemented quickly and accurately, perhaps even in a first reading. Since application of these rules is contingent on a structural analysis of the score into phrases and their gestural substructure, the present results also imply that the student pianists carried out an appropriate structural analysis, efficiently but presumably without explicit awareness. The expressive timing profile is evidence of their structural analysis.

The consistency of the students relative to the experts may have been overestimated slightly because the expert timing profiles contained human measurement error whereas the students' MIDI data did not; moreover, in most subsequent comparisons the student timing data were further stabilized by averaging over three performances, which reduced random "motor error" and brought out more clearly the pianists' intentions. While this may have tilted the comparison in favor of the students at the "high end" of the continuum, it cannot account for the large differences in consistency at the "low end." They can only be explained by assuming that some expert pianists did not wish to be highly consistent. Their intention must have been to vary their timing of repeated or similar material, and this is quite in line with what many artists say about their performance strategies. The students were much less prone to such strategies, presumably because their limited preparation (or possibly their limited experience or smaller artistic imagination) did not allow them to include multiple strategies in their performance plans. Their plans were more rigid and circumscribed; they were also safer. The experts' greater intra-individual variability carried a certain risk with it: The more different timing patterns are tried out, the more likely it is that one or the other will strike the listener as odd or mannered (see Repp, 1992a).

The students were not only as consistent as the more consistent experts in their timing, but they were equally matched in their ability to shape a *ritardando*. This ability was assessed at four different places in the music (one phrase-internal, three phrase-final), and no student consistently failed the litmus test of the parabolic curve fit. Only one pianist (P4) produced a somewhat awkward ascent to the melodic peak, though her phrase-final *ritardandi* were quite normal. Again, it was the experts who sometimes had different ideas about the shaping of these local gestures, not all of them convincing to this listener. It may still be controversial to take a curve fit to a sequence of IOIs as a measure of temporal shaping skill, but there is increasing evidence that a certain manner of changing local tempo is generally adhered to by performers and is also perceived as optimal by listeners (Sundberg and Verrillo, 1980; Repp, 1992b). This manner seems best characterized by a quadratic (or possibly cubic) function of score position (in terms of IOIs), or equivalently by a linear change in velocity (position as a function of time), all being allusions to biological motion in space (Truslit, 1938; Kronman and Sundberg, 1987; Todd, 1985, 1992, 1995, submitted; Feldman, Epstein, and Richards, 1992; Repp, 1992a, 1993; Epstein, 1995). Although an artist always has the option of deliberately deviating from such a "natural" form, (s)he does it at the (perhaps well-considered) risk of being perceived as anomalous.

The results discussed so far demonstrate that the student pianists, despite the unfavorable circumstances in which they had to play, exhibited considerable agogic skill and taste. It would not have been surprising, however, to find that their performance plans—and the structural analysis they reflect—were less detailed and somewhat impoverished compared to the experts'. It was impressive, therefore, to find that, *on the average*, the students' and experts' timing profiles were virtually the same. While some small quantitative differences existed, together with some average differences in basic tempo choices, there were no qualitative differences at all between the shapes of the respective average timing profiles. Even though the average profile is a statistical construct and not a real performance, it is a representation of significant commonality among performances and hence of a common standard or norm.[22] From this perspective, it is significant that the average profiles of students and experts were so highly similar. The finding points to a shared standard of expressive timing for this particular music, and hence also to a shared structural analysis. While there may be innumerable ways of deviating from the norm—in fact, the norm may never be realized in any particular performance—it nevertheless serves as a guiding force that "pulls" performers towards some center. This center is not pre-defined but probably has evolved through the history of performance, both of Romantic pieces in general and of "Träumerei" in particular, and it may keep changing. Precisely such an "evolutionary" theory of performance standards was recently proposed by Bowen (1993). What the present results demonstrate is that, despite differences in age, generation, and year of recording, today's student pianists seem to share the same performance standard as the very heterogeneous group of older expert pianists. This may indicate that the performance standard for "Träumerei," at least, has not changed much in recent decades. Analysis of the expert data has not revealed any obvious historical trends (Repp, 1992a).

The most important and convincing result of the present study concerns the one way in which the students differed from the experts: By several measures, but particularly in terms of the temporal organization of the intricate descent from the melodic peak in each phrase, the students' timing profiles were much more homogeneous than the experts'. The experts, even though they seemed to adhere to the same abstract standard as the students, felt much more free to deviate from it and, in doing so, showed greater individuality (and, occasionally, eccentricity) than the students. The students' individuality was a more limited and cautious one; the students seemed more strongly constrained by their common standard than many of the experts. Again, it is Bowen (1993) who has formulated a pertinent model which he in turn credits to the Russian literary critic Bakhtin (1981). Bakhtin spoke of "centripetal" and "centrifugal" forces in the everyday and artistic use of language; in Bowen's paraphrase, "the one [tends] toward unity and the need to understand each other, and the other toward the specific and the desire to express our uniqueness.... This dichotomy can also be expressed as the tension between individual expression and communication or between innovation and tradition" (Bowen, 1993, p. 143). The expert pianists, therefore, were more innovative than the students; or, more precisely, they included a number of innovative artists, for some of them were quite traditional in outlook (as far as "Träumerei" is concerned), perhaps deliberately so. Music performance thus

seems to be comparable to composition: It is generally agreed that the greatest composers, of past centuries at least, deviated in many ways from the then current compositional standards, which may have been followed religiously by lesser contemporaries and particularly by composition students.

This difference between experts and students does not come as a surprise, of course. It is its rigorous demonstration by means of objective performance analysis that is novel and deserves attention. It may be asked, however, whether the students would have produced more diverse performances if they had had the opportunity to study and rehearse the piece more carefully before the recording was made. We enter the realm of speculation here, but a negative answer seems likely. Music critics and other observers of the contemporary classical music scene often comment on the relative loss of diversity in performance, and the author can confirm this impression on the basis of having heard the present student pianists (as well as many others) in recital, playing carefully prepared programmes. One component that may contribute to the reduced originality of young artists is the competitive nature of the music business today. Music competitions, by their very nature, discourage deviation from the norm because the jury decides by consensus, and the consensus most often *is* the norm.[23] The training of today's young pianists, whether or not they have the talent to capture a top prize, is oriented towards making them successful competitors, not unique individuals. Their teachers probably assume that individuality will emerge spontaneously, and indeed it does; however, the range of the resulting individual variety is relatively restricted.

There are many other components that contribute to this phenomenon of relative uniformity among young performing artists today: The universal availability of many note-perfect recordings of the standard repertoire, which has raised expectations of technical accuracy enormously, to the detriment of interpretive originality; the increasing uniformity of these recordings as more and more young artists enter the Schwann catalogue while historical recordings fade into the background; the lack of originality in popular classical "mainstream" artists who serve as role models; the disappearance of national and regional performance traditions; the enormous influx of highly competent musicians from countries without any performance traditions in Western music; the increasing remoteness of the cultural and historical contexts that gave rise to the masterpieces that constitute the standard repertoire; and the lack of incisive life experiences in an increasingly uniform and commercialized world. It must also be remembered, however, that student pianists obviously differ from expert pianists in age and experience. It is possible that individuality increases with age and experience, and if so, there is little reason for concern. Did the great individualistic pianists of the past, such as Cortot and Horowitz, play more conventionally when they were young? This would be an interesting topic for further investigation, as would be a longitudinal follow-up study of the student pianists of today.

While the students' relatively conservative interpretations may lack the aesthetic refinement of great artists' renditions, they are interesting in their own right. Precisely because they do *not* stray too far from a common standard—because they are played as "correctly" as possible—they define that standard more precisely. Efforts to understand and model the basic principles of expressive timing would best start with prototypical profiles, leaving the modelling of originality to a later stage. Student performances are also much easier to obtain than performances of famous concert pianists in MIDI format. Furthermore, student performances may provide important information about the nature and origin of individuality in expression. For despite their relative homogeneity, the student pianists each had their own individual timing pattern, replicable (within the limits of motor control) only by themselves. These timing patterns may represent the interaction of a common structural interpretation and a common set of implicit performance rules with an individual organism whose cognitive and kinematic parameters determine the precise surface pattern of a performance. In the students' case, this interaction may be relatively uncontaminated by explicit desires to differ from the norm; the individual differences may be obligatory, as it were. Perhaps there is a relatively small set of parameters that, once determined, can predict individual variations in timing patterns and can serve as a characterization of an artist's personality. Such a parameterization of individuality—an explanation of the unexplained variance among performances—remains a project for the distant future.

REFERENCES

Bakhtin, M. M. (1981). Discourse in the novel. In M. Holquist (Ed.), *The dialogic imagination: Four Essays by M. M. Bakhtin* (C. Emerson and M. Holquist, translators). Austin, TX: U. of Texas Press.

Bowen, J. A. (1993). The history of remembered innovation: Tradition and its role in the relationship between musical works and their performances. *Journal of Musicology, 11,* 139-173.

Clarke, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception, 10,* 317-341.

Desain, P., & Honing, H. (1992). Tempo curves considered harmful. In P. Desain & H. Honing (Ed.), *Music, mind and machine* (pp. 25-40). (Thesis Publishers, Amsterdam).

Epstein, D. (1995). *Shaping time: Music, the brain, and performance* (Schirmer Books, New York).

Feldman, J., Epstein, D., & Richards, W. (1992). Force dynamics of tempo change in music. *Music Perception, 10,* 185-203.

Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A major (K.331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-103). (Royal Swedish Academy of Music, Stockholm).

Horowitz, J. (1990). *The ivory trade: Music and the business of music at the Van Cliburn international piano competition* (Summit Books, New York).

Kronman, U., & Sundberg, J. (1987). Is the musical ritard an allusion to physical motion? In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 57-68). (Royal Swedish Academy of Music, Stockholm) .

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 331-346.

Repp, B. H. (1992a). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's 'Träumerei.' *Journal of the Acoustical Society of America, 92,* 2546-2566.

Repp, B. H. (1992b). A constraint on the expressive timing of a melodic gesture: Evidence from performance and aesthetic judgment. *Music Perception, 10,* 221-242.

Repp, B. H. (1993). Music as motion: A synopsis of Alexander Truslit's "Gestaltung und Bewegung in der Musik" (1938). *Psychology of Music, 21,* 48-72.

Repp, B. H. (1994a). On determining the basic tempo of an expressive music performance. *Psychology of Music, 22,* 157-167.

Repp, B. H. (1994b). Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study. *Psychological Research, 56,* 269-284.

Repp, B. H. (submitted). Now you hear them, now you don't: An objective analysis of pianists' pitch errors.

Sundberg, J., & Verrillo, V. (1980). On the anatomy of the retard: A study of timing in music. *Journal of the Acoustical Society of American, 68,* 772-779.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3,* 33-58.

Todd, N. P. McA. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91,* 3540-3550.

Todd, N. P. McA. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America, 97,* 1940-1949.

Todd, N. P. McA. (submitted). Time and space in music performance.

Truslit, A. (1938). *Gestaltung und Bewegung in der Musik.* (Chr. Friedrich Vieweg, Berlin-Lichterfelde).

FOOTNOTES

*Journal of the Acoustical Society of America, in press.

[1] As the term is used here, expressive ("horizontal") timing does not include the asynchronies among nominally simultaneous tone onsets ("vertical" timing), which are an order of magnitude smaller than the IOIs considered here. In any chord, the onset of the highest tone (usually part of the principal melody and played by the pianist's right hand) is taken as the reference for computing IOIs.

[2] There are other ways of representing timing information—as tempo curves (see Desain and Honing, 1992), as cumulative functions of real time (Todd, 1994, 1995), or as percentage deviations from the average IOI (Gabrielsson, 1987; Palmer, 1989)—but normalized IOIs as a function of score position remain this author's preference.

[3] This is a departure from Repp (1992a), where only bar numbers and eighth-note numbers were used: "15-3-2" was "15-6" there.

[4] The performances did contain some inaccuracies—a few wrong notes and a larger number of missing and extra notes—nearly all of which were in secondary voices. (See Repp, submitted, for an error analysis.) There was no evidence that errors affected expressive timing; all performances were fluent and without hesitations.

[5] There was no evidence for a historical trend towards faster tempi in the expert data (the oldest recordings, by Davies and Cortot, were among the fastest), nor did the age at which the experts were recorded seem to be related to their tempo choices.

[6] There was no evidence for a historical trend towards faster tempi in the expert data (the oldest recordings, by Davies and Cortot, were among the fastest), nor did the age at which the experts were recorded seem to be related to their tempo choices.

[7] Such high precision should not be equated with performance quality, however; the quality of a timing profile is probably unrelated to its replicability.

[8] These correlations, like the present statistics, were computed on the subdivided, untransformed IOIs.

[9] Interestingly, the less consistent experts (Argerich, Bunin, Cortot, Ortiz, Schnabel) were those who had somewhat unusual timing patterns to begin with (see Repp, 1992a, and below). Of course, it is not known whether their inconsistency was deliberate or happenstance. It could be that less typical timing patterns are less replicable in principle (see Clarke, 1993).

[10] This figure is similar to Figure 3 in Repp (1992a), but the average profile here represents arithmetic rather than geometric means, and the abscissa has been relabeled in terms of beats rather than eighth notes.

[11] Only one of Cortot's and Horowitz's three performances was allowed among the three.

[12] While Vladimir Ashkenazy is world-famous and the late Yakov Zak was a well-known teacher in the USSR, the author has no information at all about Sylvia Capova.

[13] Repp (1992a), following the terminology of the BMDP statistical software manual, considered PCA a species of factor analysis and referred to components as "factors." The author now prefers to talk about "components," in accord with the SYSTAT software manual, but the technique is the same.

[14] Each student was represented by the average timing profile of his or her three (or two) performances, but the three performances of Cortot and Horowitz were kept separate. In each performance, the IOIs of the two renditions of bars 1-8 were averaged, and long IOIs were represented as multiple eighth-note IOIs. Thus each complete performance contained 190 IOIs. No transformation was applied. Because of the different data format, the analysis of the expert data yielded results that differed in some details from those reported in Repp (1992a).

[15] The first ratio was described somewhat awkwardly as A/(B+C) with subsequent normalization; it is equivalent to the 5A/(2B+3C) ratio reported here.

[16] Goodness of fit values (r^2) were not reported in Repp (1992a) but were computed for the present comparison.

[17]Unlike the normalized IOIs (A, B, C) used in the ratios of Table 3, these fractions of a single IOI are not normalized.

[18]The ratios for the five instances of the grace note passage in the music (the two renditions of bars 1–8 were treated individually here) were averaged. Individual pianists were fairly consistent across the five instances, and they all differed from each other in their precise grace note timing patterns, which is additional evidence for stable individual differences, despite relative homogeneity overall.

[19]The present Components II, III, IV, V, and VI correspond to Factors II, I, III, V, and IV there, and the resemblance is close. Only the marginally significant Component VII differs from the previous Factor VI.

[20]It is not in bold face in the table because the highest loading of that pianist (as well as of others at the bottom of the table) probably was on a factor that did not reach significance in the analysis.

[21]It is not in bold face in the table because the highest loading of that pianist (as well as of others at the bottom of the table) probably was on a factor that did not reach significance in the analysis.

[22]In a recent study (Repp, in preparation), the average timing profile was synthesized and presented to listeners for aesthetic judgment. It was found to be perfectly acceptable but lacking in individuality.

[23]A perceptive analysis of piano competitions is provided by Joseph Horowitz in his book, *The Ivory Trade*. Here is how he described the winner of the 1989 Van Cliburn International Piano Competition: "For one thing, he chose repertoire to highlight what he played least controversially. ...He readied his pieces not toward spontaneous, inspirational performances but toward performances that would leave nothing to chance, even under abnormal pressure. His only goal was to win" (Horowitz, 1990, pp. 101-102).

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 161-171

Quantitative Effects of Global Tempo on Expressive Timing in Music Performance: Some Perceptual Evidence*

Bruno H. Repp

This study examines the question of whether global tempo and expressive timing microstructure are independent in the aesthetic judgment of music performance. Measurements of tone interonset intervals in pianists' performances of pieces by Schumann ("Träumerei") and Debussy ("La fille aux cheveux de lin") at three different tempi show a tendency toward reduced (relative) expressive timing variation at both faster and slower tempi, relative to the pianist's original tempo. However, this could reflect merely the pianists' discomfort when playing at an unfamiliar tempo. Therefore, a perceptual approach was taken here. Experimental stimuli were created artificially by independently manipulating global tempo (3 levels) and "relative modulation depth" of expressive timing (RMD, 5 levels) in MIDI-recorded complete performances of the Schumann and Debussy pieces. Skilled pianists rated the quality of the resulting two sets of 15 performances on a 10-point scale. The question was whether the same RMD would receive the highest rating at all three tempi, or whether an interaction would emerge, such that different RMDs are preferred at different tempi. A small but significant interaction was obtained for both pieces, indicating that the listeners preferred a reduced RMD when the tempo was increased, but the same or a larger RMD when the tempo was decreased. Thus, they associated an increase in tempo with a decrease in (relative) expressive timing variation, which, in general agreement with the performance data, suggests nonindependence of the two temporal dimensions.

INTRODUCTION

Whether and how certain object or event properties remain physically and/or perceptually invariant under various kinds of transformation is an important theoretical issue that pervades psychological research (see, e.g., Warren & Shaw, 1985; Perkell & Klatt, 1986; Heuer, 1991). The psychology of music is no exception (Hulse, Takeuchi, & Braaten, 1992). For example, it is well known that musical pitch intervals and melodies remain invariant under transformations of register (i.e., transposition); that is, they both retain their frequency ratios in performance and are perceived as constituting the same melody. Some other forms of invariance in music are less well established or are in doubt. Thus, although it may seem that rhythm should scale proportionally and remain perceptually invariant across changes in global tempo—and certainly the relative note values of simple rhythms can be reproduced and recognized across changes in tempo—, several studies have suggested that subjective rhythmic organization changes with tempo (Handel & Lawson, 1983; Monahan & Hirsh, 1990; Handel, 1992; Parncutt, 1994), so that rhythms may not be executed in exactly the same way at different tempi and listeners can find it difficult to match or recognize proportionally scaled rhythmic patterns when the tempo is changed substantially (Sorkin & Montgomery, 1991; Handel, 1993).

The present study is concerned with the relative invariance or noninvariance of expressive *timing microstructure* across global tempo changes in music performance. Timing microstructure consists of continuous modulations of the local tempo, resulting in unequal intervals between

This research was supported by NIH grant MH-51230. I am grateful to Charles Nichols and Linda Popovic for extensive assistance, to Jonathan Berger (director, Yale Center of Studies in Music Technology) for allowing me to use his facilities, to the participating pianists for their patience, and to Jamshed Bharucha, Stephen Handel, Henkjan Honing, Caroline Palmer, and Burt Rosner for helpful comments on an earlier draft.

successive tone onsets, even if the corresponding notes have the same value in the score.[1] In the absence of timing microstructure, these tone interonset intervals (IOIs) would be identical; in an expressive performance, however, IOIs vary considerably and lawfully in a fashion determined both by the musical structure and the performer's interpretation and individuality (see, e.g., Gabrielsson, 1987; Palmer, 1989; Repp, 1992). *Relational invariance* (also called proportional scaling, ratiomorphic timing, or the homothetic principle; see Gentner, 1987; Heuer, 1991; Viviani & Laissard, 1991) is a key concept in research on timing control in skilled motor performance, though the tasks investigated have rarely been as complex as music performance. Relational invariance has been interpreted as evidence for a "generalized motor program" (Schmidt, 1975) that adjusts to tempo variations by means of a multiplicative rate parameter. Applied to expressive timing in music, this hypothesis implies that a change in global tempo results in a uniform compression or expansion of the timing pattern, leaving the ratios of successive IOIs constant.

Clarke (1982), following up on earlier observations by Michon (1974), has suggested that timing microstructure does not remain relationally invariant as the tempo of a performance changes, due to rhythmic reorganization. Although Clarke's data are weak (see Repp, 1994a), his discussion is reasonable: If global tempo changes are large enough to cause rhythmic reorganization, then this will probably be reflected in expressive timing and cause deviations from relational invariance. Desain and Honing (1994) presented stronger evidence for noninvariance of timing microstructure in a piano piece played at three different tempi. However, they did not make clear what, if any, systematic principle underlay the deviations from invariance. Some of the clearest deviations occurred in the very brief IOIs associated with grace notes, where some perceptual or motoric lower limit may have been reached that prevented proportional scaling. The present study, in contrast, focused on the timing microstructure of relatively long IOIs, where such limits presumably do not play any role. More importantly, the present study was concerned with a relatively limited range of global tempi, over which a performer might reasonably be expected to maintain a particular rhythmic organization. Thus, while better demonstrations of *qualitative* changes in expressive timing as a function of changes in global tempo are needed, the research reported here focuses on the orthogonal dimension of *quantitative* changes in timing microstructure.

A qualitative change is one that affects the shape of the timing profile (IOI duration as a function of metrical position) and thus suggests a change in the underlying rhythmic structuring (e.g., a peak appears in the timing profile where previously there was none, or a peak disappears completely while other features remain relatively constant), whereas a quantitative change affects the magnitude of all peaks and valleys in the profile, suggesting relatively understated or exaggerated expression of the same underlying organization. Different degrees along this quantitative continuum are captured by the concept of *relative modulation depth* (RMD), which will be defined more precisely below. Note that *absolute* modulation depth (i.e., the absolute range of variation of the IOI durations) is likely to increase as global tempo decreases (i.e., as the average IOI duration increases). However, if relational invariance holds, then the variation increases simply by a multiplicative factor and the RMD remains constant. The question addressed in the present research, then, was whether the RMD does in fact remain constant when the tempo of a performance is changed. A less formal way of posing this question is: Does a change in overall tempo affect the degree of expressiveness of a performance (in the timing domain)?

Some Relevant Performance Data

Repp (1994a) investigated whether relational invariance of expressive timing microstructure held in two pianists' performances of Robert Schumann's famous miniature, "Träumerei," played at three different tempi. Even though the timing profiles of all performances by the same pianist were highly similar, statistical analysis did show significant deviations from relational invariance. A subsequent regression analysis revealed systematic trends in these deviations, which are illustrated in Figure 1.[2]

This figure plots the logarithms of the IOIs (n = 254) at the original preferred (medium) tempo against the logarithms of the same IOIs at faster and slower tempi, for each of the two pianists (LPH and BHR). To reduce random variability, the IOI durations were averaged over three performances at the same nominal tempo before the logarithms were computed. All IOIs are nominally eighth notes, so that all variability is due to expressive timing alone. (Notationally longer IOIs in the music were subdivided into equal parts cor-

responding to eighth-note IOIs.) If relational invariance holds, then corresponding IOIs at two different tempi (T1, T2) should be proportional, so that $IOI_{T2} = \varsigma IOI_{T1}$, where ς is a constant representing the tempo change. It follows that $\log(IOI_{T2}) = \log(\varsigma) + \log(IOI_{T1})$; that is, the relationship between the logarithms of the IOIs should be linear with a slope of 1 (parallel to the solid diagonal in Figure 1) and the intercept $\log(\varsigma)$. However, the slopes of the dotted lines fit to the data in Figure 1 are significantly less than 1, more so for LPH (Figure 1a), a professional pianist, than for BHR (Figure 1b), an amateur.[3]

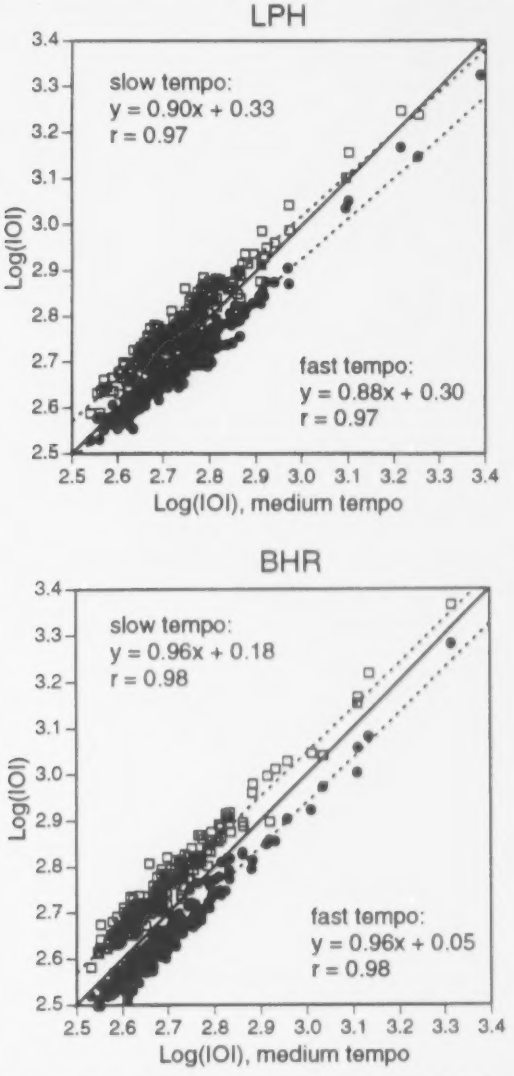


Figure 1. **Relationships between logarithms of IOIs at a medium tempo and at a fast and slow tempo, respectively, in performances of Schumann's "Träumerei" by two pianists (LPH and BHR). Data from Repp (1994a).**

This alone would not be sufficient evidence for a deviation from proportionality because the slope of a regression line declines in proportion to the correlation between two variables. The relevant evidence is the fact that the slopes in Figure 1 are smaller than the correlations, implying that the standard deviations of the log(IOI) values were reduced at slow and fast tempi relative to medium tempo. (The slope divided by the correlation equals the ratio of the standard deviations.) In other words, when the tempo was increased, long IOIs decreased proportionally more than did short IOIs; when the tempo was decreased, long IOIs increased proportionally less than did short IOIs. In each instance, a change of tempo resulted in a reduced RMD, or a compression of the timing profile relative to the original tempo. The measure of RMD is the slope of the regression line divided by the correlation.

These observations suggest that expressive timing is not independent of global tempo. However, the nature of this dependence is surprising: Instead of there being a unidirectional change in "expressiveness" with tempo (with a relative decrease in expressiveness as the tempo gets faster being much more plausible than the reverse), it seems that the two pianists, especially LPH, played less expressively at both slow and fast tempi.[4] This could have a relatively trivial explanation, however: When a musician is asked to play at a tempo other than the preferred one, (s)he may feel less comfortable and/or may have to devote some attention to sticking to the prescribed tempo, which then may result in a restricted RMD.[5] Alternatively, it could be that there is an optimal (medium) tempo for a piece of music, which somehow permits the greatest expressive freedom. Either of these possibilities may be called an "optimal tempo" hypothesis. The alternative hypothesis, which is not supported by the performance data, may be called "the faster, the less expressive." The null hypothesis, of course, is that relational invariance holds.

A Perceptual Approach

In order to circumvent possible artifacts due to performers' tempo preferences, the present study took a perceptual approach. If RMD tends to vary with global tempo in performance, then musicians should have corresponding expectations as listeners. These expectations were assessed here by varying global tempo and RMD *independently* in a performance, and by asking skilled pianists to evaluate the quality of the resulting versions. Given that a particular RMD is preferred at the medium (original) tempo, then it may be asked

whether the same RMD is also preferred at a slower or faster tempo. Such a result would be consistent with the null hypothesis that expressive timing is relationally invariant across tempo changes. In that case, the variations in RMD observed in the "Träumerei" performances could be regarded as artifacts related to pianists' tempo preferences. Alternatively, the pianist judges might prefer a larger RMD at the slow tempo and a smaller RMD at the fast tempo. This would support the hypothesis of "the faster, the less expressive" and would suggest that only the reduction in RMD at a slow tempo is a performance artifact. Finally, it is conceivable that the perceptual judgments will mirror the performance data, with a smaller RMD being preferring at both a slow and a fast tempo. This would support the "optimal tempo" hypothesis. Statistically speaking, both of the latter findings represent an *interaction* between global tempo and RMD, whereas the null hypothesis predicts no interaction.

In his earlier related study, Repp (1994a) focused on performance measurements but also included a perceptual test. He took the first 8 bars of each of the two pianists' "Träumerei" performances at the three tempi and changed their tempo artificially by stretching or shrinking all IOIs proportionally, so that RMD remained constant. These modified performances were then paired with unmodified performances having the same global tempo (i.e., the same total duration), and pianist listeners were asked to indicate which performance in each pair was the original one—that is, which sounded more natural. This was a very difficult task, and accuracy was barely above chance level. These results were interpreted as supporting the hypothesis of relational invariance: Changing the tempo while keeping RMD constant did not seem to result in a noticeable deterioration of expressive quality. The present study went two steps further by varying both global tempo *and* RMD orthogonally in *complete* performances of piano pieces.

Experiment 1 was preceded by a very similar experiment that yielded unclear results and therefore will not be reported in detail. It differed from the experiment reported below in that both global tempo and RMD varied over a smaller range, which probably made the stimuli too difficult to discriminate and judge reliably. It also presented the listeners with integral performances rather than with 8-bar excerpts, as described below. Nine pianists participated as listeners, one of them being LPH, whose medium-tempo performance had formed the basis of the experimental materials. The single remarkable result was that LPH was the only listener who showed a striking interaction of global tempo and RMD in her ratings, which fit the pattern of "the faster, the less expressive."[6] LPH was the most experienced pianist among the listeners, but she may also have been specially attuned to her own expressive microstructure. While not much could be concluded from this intriguing finding, it gave rise to the hope that more consistent results from a group of pianist judges might be obtained when global tempo and RMD were varied over a wider range.

EXPERIMENT 1

Method

Listeners

Ten pianists participated as listeners. Seven of them were graduate students of piano at the Yale School of Music, one was a graduate student of music theory, one was an undergraduate, and one was a serious amateur (the author).[7]

Materials

Fifteen complete versions of Schumann's "Träumerei" were generated by transforming a single original performance by LPH. That performance was one of three recorded at LPH's preferred ("medium") tempo on a Roland RD-250s digital piano with DP-2 pedal switch (see Repp, 1994a, 1994b, for details). The performance was technically accurate and had fine artistic expression; despite the slightly synthetic sound ("Piano 1"), it was a pleasure to listen to. The performance data were stored in MIDI format (note onsets and offsets, velocities, and pedal onsets and offsets).

The transformation method involved several steps and decisions. Following Repp's (1992, 1994a) methods of data analysis, eighth-note IOIs were derived from the onsets of the tones with the highest pitch in each cluster of nominally simultaneous tones. IOIs nominally longer than one eighth note were divided into eighth-note IOIs of equal length, to be added up again after transformation. This part was straightforward. The tricky question was how to deal with onset asynchronies, grace notes, note offsets, and pedal information when manipulating tempo and RMD.

Repp (1994a), in his analyses of LPH's and BHR's performances of "Träumerei," has provided some evidence suggesting that onset asynchronies in chords and overlap times (degree of *legato*) of successive tones do not change systematically with changes in global tempo. However, since it

proved technically cumbersome to keep these small intervals constant while transforming the primary IOIs, the MIDI score of LPH's original performance was first edited to eliminate all onset asynchronies and overlaps. Thus, all notes with nominally simultaneous onsets were made to start at the same time as the note from which the IOI was computed, viz., the one with the highest pitch. Similarly, all note offsets in the MIDI score were "regularized" by making them coincide with following note onsets, according to their nominal value in the score. This eliminated overlaps between successive *legato* notes (the original performance was almost entirely *legato*) as well as gaps between notes played *non-legato*, such as repeated notes of the same pitch.[8] Subsequent listening suggested that the performance did not suffer in expressive quality from these manipulations. Since pedaling was almost continuous and created extensive acoustic tone overlaps (see Repp, 1995b), the elimination of overlaps and gaps in the MIDI score had few audible consequences.

After this regularization, the remaining MIDI events that did not coincide with tone onsets were a few grace notes and the ubiquitous pedal onsets and offsets. According to earlier analyses, grace note timing in this music was relationally invariant and pedal timing often changed with tempo, although it did not always exhibit relational invariance (Repp, 1994a, 1995b). It was decided to keep the timing of both these events relationally invariant. That is, after transforming the IOIs, the grace note onsets and pedal events were moved so that they remained in the same *relative* temporal position within the IOI in which they occurred.

The eighth-note IOIs themselves were transformed by first computing their natural logarithms, then multiplying them by a constant $\underline{b}$ and adding a constant $\underline{a}$, and finally taking the antilogarithm of the result.[9] The values of $\underline{b}$ (the measure of RMD) were chosen to be 0.6, 0.8, 1.0, 1.2, and 1.4, where a coefficient of less than 1 represents a compression and a coefficient larger than 1 represents an expansion of RMD compared to the original performance (cf. Figure 1). Starting with the original medium-tempo performance ($\underline{a}$ = 0, $\underline{b}$ = 1), two values of the intercept $\underline{a}$ were chosen to generate slower and faster versions whose tempo still seemed aesthetically acceptable. The total durations of the resulting fast, medium, and slow performances differed by increments of 25%. A different $\underline{b}$ coefficient was then applied to the medium-tempo performance, and an accompanying value of $\underline{a}$ was found by trial and error, so as to keep the total duration constant. The $\underline{a}$ coefficients for the remaining versions could then be determined arithmetically, as they were a simple linear function of the $\underline{b}$ coefficients. All these coefficients are shown in Table 1.

Table 1. *Additive ($\underline{a}$) coefficients used in Experiment 1. These are the intercepts of the linear functions relating the original and transformed ln(IOI) values, with the $\underline{b}$ coefficients being their slopes.*

Tempo	$\underline{b}$ coefficient					Duration
	0.6	0.8	1.0	1.2	1.4	(s)
Fast	2.319	1.046	-0.227	-1.504	-2.781	115.0
Medium	2.546	1.273	0.000	-1.277	-2.554	144.3
Slow	2.767	1.494	0.221	-1.056	-2.333	180.0

All data manipulations were carried out in a spreadsheet/graphics program (DeltaGraph Professional) into which the original MIDI data had been imported as text files. After transformation, the IOIs were cumulated back into absolute onset times, and the data were reconverted into MIDI files for audio output via the Roland RD-250s digital piano. To get multiple ratings of each version, the 24-bar piece was divided into three 8-bar sections (excerpts) that were presented and evaluated separately.[10]

Procedure

The 45 8-bar sections were recorded onto DAT tape, together with the complete original regularized medium-tempo performance, which served as familiarization. All listeners were tested individually in a quiet room and heard the music over Sennheiser HD540II earphones. The DAT recorder was programmed to deliver the excerpts in a different order to each listener.

Each listener first heard the complete performance. (S)he was asked to consider it approximately 7 or 8 on a 10-point scale (1 = poor, 10 = excellent) and to judge the quality of the following performances relative to it, as well as relative to any other preceding versions of the same tempo.[11] The test excerpts were presented in three groups of 15, corresponding to the three 8-bar sections of the piece, which were presented in the same natural order to all subjects (i.e., first bars 1-8, then bars 9-16, and finally bars 17-24). Each group was divided into three blocks of five versions each, with the global tempo being constant within each block. The order of blocks (global tempi) within groups, and the order of excerpts (RMD values) within each block, were variable and approxi-

mately counterbalanced across listeners. Each individual listener, however, received the same order of global tempi within each group, and the same order of RMD values within tempi in any given group (but different orders in different groups), in order to equate any sequential context effects across tempi.

The listeners were asked to assign a rating at the end of each performance excerpt. They were asked to try to use the whole range of the scale and to avoid giving the same rating to two excerpts in the same block; decimals and ratings outside the 10-point range were permitted (but rarely used). It was emphasized that discrimination among the excerpts within a block was much more important than the relative ratings of the performances according to tempo (i.e., between blocks). The listeners were thus asked specifically to focus on the RMD dimension and to indicate their relative preference among excerpts having the same global tempo. Each group of 15 excerpts was followed by a short break.

Results and Discussion

The average ratings of the 10 judges are shown in Figure 2 as a function of RMD (i.e., the b coefficient), separately for the three global tempi. As expected, at the medium tempo the highest rating was given to the original (regularized) performance ($b = 1$). At the fast tempo, however, the highest rating was given to the performance with $b = 0.8$, whereas at the slow tempo it was given to the performance with $b = 1.2$. The tempo by RMD interaction was significant in a repeated-measures ANOVA [$F(8,72) = 3.41$, $p < .003$]. In addition, there was a significant main effect of RMD [$F(4,36) = 4.81$, $p < .004$]. The apparent preference for the medium tempo over the other tempi was not reliable due to large individual variability, and there were no differences or interactions due to the three 8-bar sections of the piece. Pairwise comparisons of tempi suggested that the interaction with RMD was reliable for slow vs. fast [$F(4,36) = 4.09$, $p < .008$] and for slow vs. medium tempo [$F(4,36) = 3.33$, $p < .03$], but not quite for medium vs. fast tempo [$F(4,36) = 2.32$, $p < .08$].

These results support the hypothesis of "the faster, the less expressive": Listeners exhibited a preference for a reduced RMD at a fast tempo and for an enhanced RMD at a slow tempo. The effect is small, however. It was possible to assess the reliability of each individual subject's results by considering the three 8-bar excerpts as a random factor crossed with the fixed factors of RMD and tempo. In these individual repeated-measures ANOVAs, eight of ten listeners showed significant ($p < .05$) main effects of RMD and nine showed significant main effects of tempo, which suggests that they could discriminate among the different performances. However, only three subjects showed a significant RMD by tempo interaction. Of course, these individual ANOVAs had less statistical power than the overall analysis, but they demonstrate the relative fragility of the crucial interaction.

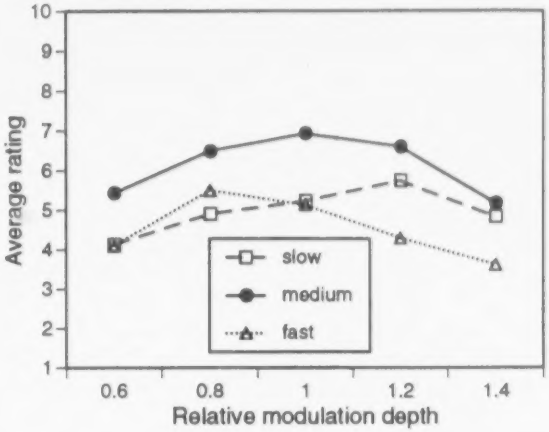


Figure 2. **Average ratings by 10 subjects of 15 performances of Schumann's "Träumerei" varying in global tempo and in RMD.**

Experiment 2 represents an attempt to replicate this interaction with a different piece of music.

EXPERIMENT 2

Some Perfroamce Data

The music in this study was "La fille aux cheveux de lin," from Book I of Debussy's preludes. The complete prelude was performed 7 times by a talented young pianist, a second-year graduate student at the Yale School of Music. The first three times she played it at her preferred tempo, alternating with three other pieces that were recorded in the same session. At the end of the session, she was asked to play the Debussy piece twice each at the slowest and fastest tempi that she found aesthetically acceptable. The instrument was a Yamaha MX100A Disclavier (an upright acoustic piano with added electronic components, connected to a microcomputer) located at the Yale Center for Studies in Music Technology. The performances were recorded in MIDI format.

The "primary" note onsets (i.e., of the note with the highest pitch in each cluster) were identified in the MIDI scores, and IOIs were calculated. The IOIs were then averaged across the two or three performances with the same nominal tempo, to

reduce random variability. In contrast to Schumann's "Träumerei," which contains mainly eighth-note IOIs, the Debussy piece contains a variety of nominal IOI durations: sixteenth notes, eighth notes, and longer notes. To eliminate the contribution of nominal differences in IOI duration and leave only expressive timing variation, all IOIs were "normalized" to sixteenth-note units by dividing longer IOIs by the number of sixteenth-note units they contained. The scatter plot of log(IOI) values comparing the different tempi is shown in Figure 3a. The deviation of the slopes of the regression lines from unity was even more striking here than in Figure 1, though there was also greater variability.[12] However, the slopes of the regression lines were clearly smaller than the correlations, indicating again a reduction in RMD at both fast and slow tempi.

Figure 3b shows a comparison of the most prevalent short IOIs, those of sixteenth and (here not normalized) eighth notes. Although both types of IOIs had regression lines with slopes of less than 1, sixteenth notes had especially shallow slopes.[13] The slopes were smaller than the correlations in all but one case (eighth notes at the fast tempo). It seems that the reduction in RMD was most pronounced for the shortest notes in the music.

These observations confirm the trends found in the "Träumerei" performances. Again, it seems that the pianist played more expressively at her preferred tempo than at either a faster or a slower tempo. The more complex inventory of IOIs raised the question, however, how the RMD transformation in the perceptual test materials should be handled: Should it be carried out on the original undivided IOIs or on the IOIs divided into sixteenth notes? Should sixteenth notes be treated differently from the others? It was decided, somewhat arbitrarily, to use the same method as in Experiment 1, viz., to divide all IOIs into sixteenth-note units before transformation and then to add up the transformed fractions to reconstitute the longer IOIs.

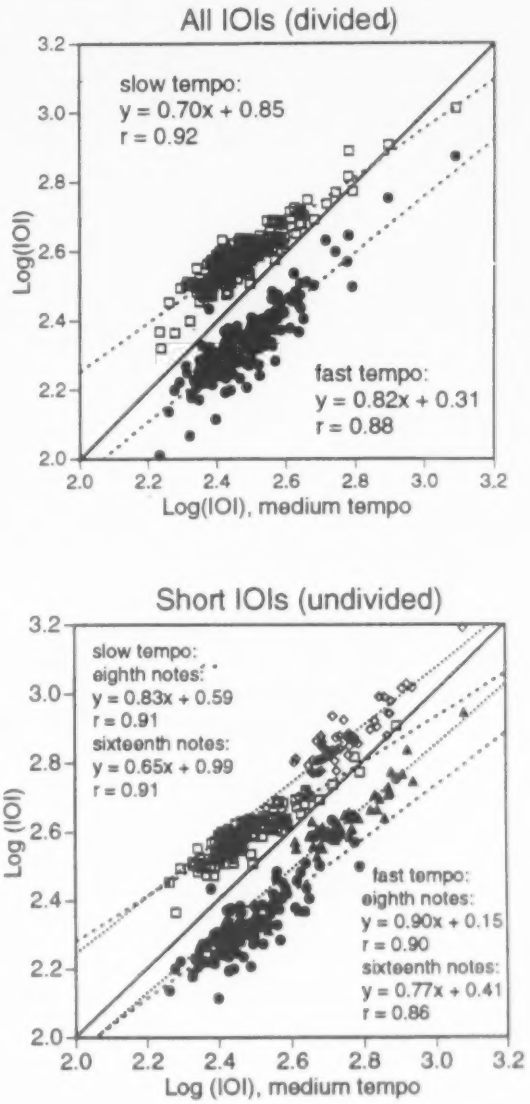


Figure 3. **Relationships between logarithms of IOIs at a medium tempo and at a fast and slow tempo, respectively, in performances of Debussy's "La fille aux cheveux de lin" by one pianist. (a) All IOIs divided into sixteenth-note units. (b) Sixteenth notes (circles and squares, dashed regression lines) and undivided eighth notes (triangles and diamonds, dotted regression lines).**

Method

Listeners

Nine pianists participated as listeners. They included four graduate students of piano at the Yale School of Music (one of whom had provided the performances just described), three excellent undergraduate pianists (all had performed as soloists with the Yale Symphony Orchestra the same season), one semi-professional accompanist, and one amateur (the author). All indicated that they knew the music well.

Materials

One of the original medium-tempo performances was selected as the basis for the experimental manipulations. This performance and all its descendants were reproduced on the Roland RD-250s digital piano used in Experiment 1, which sounded very acceptable and avoided problems connected with acoustic recording.[14] As in

Expe iment 1, the performance was then "regularized" by synchronizing all note onsets and offsets according to their notated values. In each cluster of nominally simultaneous events, the onset of the note with the highest pitch again served as the reference. Regularization did not seem to affect performance quality. Note events that were left in their original relative temporal positions within IOIs included two arpeggiated chords (bars 12 and 35), three "split" left-hand chords (bars 6, 16, 30), and two broken octaves at the end of the piece (bars 36 and 37). Pedal onsets and offsets also remained in their original relative positions. The damper pedal was used extensively throughout the music.

Transformation of the IOIs was carried out according to the same design and regime as in Experiment 1. All IOIs were divided into sixteenth-note intervals before transformation. The a and b coefficients and the overall durations of the performances are shown in Table 2. The slow and fast tempi were those chosen by the pianist herself in her slow and fast performances. While the 23% increase in duration from medium to slow tempo was comparable to that in Experiment 1, there was a larger (44%) increase here from fast to medium tempo.

Table 2. *Additive (a) coefficients used in Experiment 2. These are the intercepts of the linear functions relating the original and transformed ln(IOI) values, with the b coefficents being their slopes.*

Tempo	b coefficient					Duration
	0.6	0.8	1.0	1.2	1.4	(s)
Fast	1.931	0.784	-0.365	-1.518	-2.675	93.8
Medium	2.297	1.150	0.000	-1.153	-2.310	135.1
Slow	2.505	1.358	0.209	-0.944	-2.101	166.4

Procedure

The procedure was the same as in Experiment 1, except that the performances were presented in integral form, so that each was judged only once by each subject. The original medium-tempo performance again served as familiarization, and subjects were asked to assign it a "9" on the 10-point rating scale. It was followed by three blocks of 5 performances each. The order of blocks (global tempi) and of performances (RMD values) within blocks was varied across subjects, but the same order of RMD values was used in the three blocks for any given subject (except for the first two subjects who received different orders in the three blocks).

Results and Discussion

The average ratings are shown in Figure 4. As expected (and instructed), subjects gave a high rating to the original medium-tempo performance, but they liked the performance with slightly exaggerated timing variation (b = 1.2) just as much, and even the most exaggerated performance (b = 1.4) received a rather high rating.[15] Performances with reduced timing variation (b = 0.8, 0.6) were liked much less. The slow performances present a rather similar picture, though with lower ratings overall. At the fast tempo, however, this asymmetry was absent, and understated performances were actually rated slightly higher than exaggerated ones. This pattern of results again represents a significant RMD by tempo interaction [$F(8,64) = 3.23$, $p < .004$]. In addition, there were significant main effects of tempo [$F(2,16) = 10.29$, $p < .002$] and of RMD [$F(4,32) = 6.30$, $p < .0008$]. Pairwise comparisons of tempi showed the interaction to be highly significant for medium versus fast tempo [$F(4,32) = 6.52$, $p < .0007$], marginally significant for slow versus fast tempo [$F(4,32) = 2.97$, $p < .04$], and nonsignificant for medium versus slow tempo [$F(4,32) = 0.69$].[16]

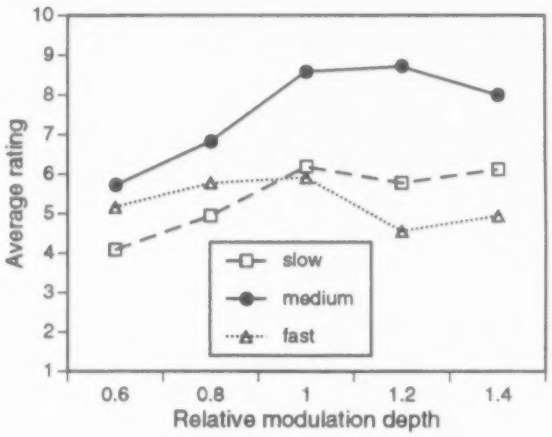


Figure 4. **Average ratings by 9 subjects of 15 performances of Debussy's "La fille aux cheveux de lin" varying in global tempo and in RMD.**

The results for medium versus fast tempo, and especially for slow versus fast tempo, are in agreement with those of Experiment 1 and thus support the hypothesis of "the faster, the less expressive." The comparison of slow and medium tempo, however, tends in the opposite direction. A significant effect in this comparison might have given support to the "optimal tempo" hypothesis, but the nonsignificant trend does not warrant any

conclusions. The fact that the tempo difference between slow and medium was smaller in this experiment than that between medium and fast may have been partially responsible for the result. Thus, on the whole, the results of Experiment 2 are consistent with those of Experiment 1, especially if only the extreme tempi are considered. Again, however, the interaction represents a relatively small effect.

GENERAL DISCUSSION

The present results suggest, tentatively, that expressive timing microstructure is not completely independent of global tempo, even when the tempo variations do not affect rhythmic organization (i.e., do not lead to interpretable qualitative changes in the timing profile). Although the timing profiles of performances played at different tempi may be highly similar (Repp, 1994a), they do exhibit statistically reliable differences (see also Desain & Honing, 1994). Correspondingly, listeners seem to expect the timing pattern to change with global tempo. This change, as long as the tempo stays within aesthetically acceptable limits, seems to be quantitative rather than qualitative in nature: It takes place along the continuum of RMD or degree of expressiveness.

There is a discrepancy, however, between the (admittedly limited) production and perception data presented here. The perceptual judgments suggest that listeners expect the RMD to be reduced at a fast tempo, and this is in agreement with the performance measurements. At a slow tempo, however, listeners seem to expect the RMD to be increased (Exp. 1) or unchanged (Exp. 2), whereas pianists appear to reduce the RMD when playing at a slow tempo. However, as already pointed out, this reduction may be due to the exigencies of playing at an unfamiliar tempo: The pianist may have to devote attention to keeping the tempo, at the expense of expression. Perhaps this effect would disappear if a pianist practiced a piece at a slow tempo. Although tempo preferences may also affect perceptual evaluation, the perceptual data came from a larger sample of pianists whose tempo preferences presumably were both diverse and less pronounced (unless a pianist had studied the piece recently). Therefore, the perceptual data may be more representative than the performance data.

The present results must be considered preliminary for a number of reasons. First, they derive from only two compositions, both relatively slow and lyrical in character; it remains to be seen whether the results generalize to other pieces. Second, the transformations were applied to only one specific performance of each piece, which may also limit the generality of the findings. Third, the RMD transformation procedure involved certain decisions that future research may call into question. While the elimination of tone onset asynchronies and overlaps is relatively uncontroversial and did not seem to harm performance quality, due in part to extensive pedal use, the treatment of longer IOIs, *arpeggi*, and grace notes is more critical. There is good evidence from earlier studies (Repp, 1994a, 1995b) that the timing of the relatively slow grace notes in "Träumerei" remains relationally invariant with changes in global tempo; however, faster grace notes may behave differently (see Desain & Honing, 1992, 1994). The timing of the *arpeggi* in the Debussy piece was allowed to vary proportionally with global tempo, though this was perhaps not the optimal procedure. Most importantly, the treatment of nominally long IOIs in terms of equal subunits is in need of a firmer empirical and theoretical basis.

The present study explored a new methodology that endeavors to stay as close as possible to genuine artistic performance and aesthetically informed listening. There have been few if any previous studies in which integral performances have been subjected to computer-controlled transformation that preserved their human quality and general aesthetic acceptability. Power-function transformations of expressive timing and the concept of RMD seem to have ecological validity in that they appear to preserve important characteristics of artistic time management while moving along a continuum from understatement to exaggeration. Desain and Honing's (1992) calculus for expressive transformations incorporates a very similar procedure. While the present approach was motivated by empirical observations, theirs was based mainly on theoretical considerations or common sense. However, their system, which makes possible much more sophisticated, structure-sensitive transformations, has not been used in formal experiments so far. The power-function transformation also seems intuitively compatible with Repp's (1992) finding of parabolic timing functions and Todd's (1992, 1995) recent model of expressive timing, which represents tempo modulations as linear changes in the velocity of musical motion over time. The MIDI-based manipulation of human performances seems a promising technique for certain purposes, as long as music performance synthesis is not sufficiently developed to produce truly human-like outcomes.

A final issue that needs to be commented on relates back to the introductory paragraph on transformational invariance. The invariance of a melody across changes in pitch register can be demonstrated by (1) asking musicians to play the same melody in a different key and measuring the resulting pitch relationships and (2) by asking listeners to identify the transposed melody as being the same as the original. The same method may be applied to a rhythmic pattern played at different tempi. When it comes to expressive microstructure, however, we are dealing with a subtle, complex, and subcategorical form of variation whose invariance across transformations is difficult to judge directly, especially when quantitative rather than qualitative differences are at stake. That is, if a musician were asked to play a piece with exactly the same degree of expressive timing but at a different tempo, (s)he would either deny that this is possible or go ahead but not really know whether (s)he is following the instructions. What a musician can do is play the same music at a different tempo with the expression that seems best at that tempo, and this is what the pianists in the present study did. Similarly, even highly experienced listeners would find it extremely difficult to judge whether two performances differing in global tempo have exactly the same degree of tempo modulation.[17] What these listeners can do is judge the expressive quality of performances varying in tempo, and this is what the present listeners were asked to do. The interdependence of tempo and RMD demonstrated here thus resides in the domain of artistic performance and aesthetic evaluation, not in that of psychophysical judgment. Therefore, the present results do *not* demonstrate that the intended or judged degree of expressive timing variation depends on tempo, but rather that the aesthetically most satisfying RMD shows such a dependency.

REFERENCES

Clarke, E. F. (1982). Timing in the performance of Erik Satie's 'Vexations.' *Acta Psychologica, 50,* 1-19.

Desain, P., & Honing, H. (1992). Towards a calculus for expressive timing in music. In P. Desain & H. Honing (Eds.), *Music, mind and machine* (pp. 175-214). Amsterdam: Thesis Publishers.

Desain, P., & Honing, H. (1994). Does expressive timing in music performance scale proportionally with tempo? *Psychological Research, 56,* 285-292.

Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A major (K. 331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-103). Stockholm: Royal Swedish Academy of Music (Publication No. 55).

Gentner, D. R. (1987). Timing of skilled motor performance: Tests of the proportional duration model. *Psychological Review, 94,* 255-276.

Handel, S. (1992). The differentiation of rhythmic structure. *Perception & Psychophysics, 52,* 497-507.

Handel, S. (1993). The effect of tempo and tone duration on rhythm discrimination. *Perception & Psychophysics, 54,* 370-382.

Handel, S., & Lawson, G. R. (1983). The contextual nature of rhythmic interpretation. *Perception & Psychophysics, 34,* 103-120.

Heuer, H. (1991). Invariant relative timing in motor-program theory. In J. Fagard & P. H. Wolff (Eds.), *The development of timing control and temporal organization in coordinated action* (pp. 37-68). Amsterdam: Elsevier.

Hulse, S. H., Takeuchi, A. H., & Braaten, R. F. (1992). Perceptual invariances in the comparative psychology of music. *Music Perception, 10,* 151-184.

Michon, J. A. (1974). Programs and "programs" for sequential patterns in motor behaviour. *Brain Research, 71,* 413-424.

Monahan, C. B., & Hirsh, I. J. (1990). Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics, 47,* 7227-242.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 331-346.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception, 11,* 409-464.

Perkell, J. S., & Klatt, D. H. (Eds.). (1986). *Invariance and variability in speech processes.* Hillsdale, NJ: Erlbaum.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America, 92,* 2546-2568.

Repp, B. H. (1994a). Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study. *Psychological Research, 56,* 269-284.

Repp, B. H. (1994b). On determining the basic tempo of an expressive music performance. *Psychology of Music, 22,* 157-167.

Repp, B. H. (1995a). Acoustics, perception, and production of *legato* articulation on the piano. *Journal of the Acoustical Society of America* (in press).

Repp, B. H. (1995b). Pedal timing and tempo in expressive piano performance. *Psychology of Music* (in press).

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review, 82,* 225-260.

Sorkin, R. D., & Montgomery, D. A. (1991). Effect of time compression and expansion on the discrimination of tonal patterns. *Journal of the Acoustical Society of America, 90,* 846-857.

Todd, N. P. McA. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91,* 3540-3550.

Todd, N. P. McA. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America, 97,* 1940-1949.

Viviani, P., & Laissard, G. (1991). Timing control in motor sequences. In J. Fagard & P. H. Wolff (Eds.), *The development of timing control and temporal organization in coordinated action* (pp. 1-36). Amsterdam: Elsevier.

Warren, W. H. Jr., & Shaw, R. E. (Eds.) (1985). Persistence and change. Proceedings of the First International Conference on Event Perception. Hillsdale, NJ: Erlbaum.

FOOTNOTES

[*] *Music Perception*, in press.

[1] Timing microstructure, as defined here, does not include other temporal aspects of performance, such as asynchronies among the onsets of nominally simultaneous tones, overlaps among

successive tones, and pedal timing. All these phenomena may be subsumed under the category of *temporal microstructure*.

[2]This figure was not included in Repp (1994a) and reflects analyses conducted after that paper went to press. See Repp (1994b) for detailed information about the pianists' tempo choices.

[3]The significance of the deviations of the slopes of the regression lines from unity was tested by computing the correlations between the log(IOI) values at the medium tempo and the differences between corresponding log(IOI) values at the fast and medium, or the slow and medium, tempi. These "difference correlations" were -0.50 and -0.43 ($p < .0001$), respectively, for LPH, and -0.18 ($p < .01$) and -0.21 ($p < .001$), respectively.

[4]Actually, the evidence for the slow tempo is somewhat ambiguous, for the following reason: There is noise in the data, as reflected in the imperfect correlations. Some or all of this noise may be due to imperfect motor control, whose variability may be independent of tempo, or nearly so. If so, then this random variance will be larger relative to the systematic variance at a fast tempo than at a slow tempo, which leads to the prediction that the total variance of log(IOI) values should decrease with tempo. This is clearly counter to the finding of a smaller variance at the fast than at the medium tempo, but it is consistent with the smaller variance at the slow than at the medium tempo and may partially account for it.

[5]The tempi were set by a metronome which was turned off before the performance began. The medium tempo, however, corresponded to each pianist's spontaneously chosen tempo in an initial performance (see Repp, 1994b, for details of procedure).

[6]Of course, this implies "the slower, the more expressive" which, paradoxically, contradicted her own performance (Figure 1a).

[7]Although the author had had prior experience with the stimuli, he was blind to their order in the test and had no bias in favor of a particular hypothesis. It may be assumed that all listeners were familiar with Schumann's "Träumerei," which is perhaps the most famous piano composition of the Romantic period.

[8]These overlaps and gaps are defined with respect to the note onsets and offsets in the MIDI score. *Acoustic* and *perceptual* overlaps and gaps are a different matter which need not be considered here (see Repp, 1995a). Despite the elimination of gaps in the MIDI score, the onset of a repeated tone remained clearly perceptible, due to the acoustic decay of the preceding tone prior to its nominal offset.

[9]This is equivalent to applying the power function $y = e^{a}x^{b}$, also suggested by Desain & Honing (1992). Natural logarithms were used here for a trivial technical reason; note that the performance data (Figure 1) are displayed in terms of base 10 logarithms.

[10]Actually, each complete performance consisted of 32 measures, as bars 1-8 were repeated. This repeat was not used in the experiment. The duration of an 8-bar excerpt at any tempo thus was roughly one fourth, not one third, of the total duration given in Table 1. Each excerpt started with an upbeat, and some slight adjustments were made in the MIDI scores to achieve smooth beginnings and endings of each excerpt when presented separately.

[11]Since an artificial transformation can hardly improve a fine artistic performance, it seemed highly likely that the original would be rated higher than any other version when it recurred as a test stimulus. Therefore, it seemed appropriate to anchor the rating scale initially by asking listeners to assign a relatively high rating to the original performance, which made it likely that similarly high ratings would be assigned to its 8-bar sections when they recurred during the test, so that most of the steps of the rating scale were available for evaluation of the transformations. Also, to observe a shift of the peak rating as a function of global tempo, it was desirable that there be a clear central peak (i.e., at $b = 1$) in the ratings of the five medium-tempo versions.

[12]The "difference correlations" (-0.70 and -0.39 for medium vs. slow and medium vs. fast tempo, respectively) were highly significant.

[13]All difference correlations were significant, reaching a remarkable -0.77 for sixteenth notes at the slow tempo.

[14]The only necessary modification was elimination of the soft pedal, which caused unpleasantly abrupt changes in dynamics on the Roland. An apparent misreading by the pianist of three eighth notes as sixteenth notes was also corrected at this stage.

[15]Unlike pianist LPH, who seemed exceptionally sensitive to modifications of her own timing microstructure in the earlier version of Experiment 1, the pianist who had provided the original performance for this experiment did not give ratings that were radically different from those of the other subjects.

[16]Since individual subjects gave only a single rating of each performance, individual results could not be analyzed statistically and were somewhat variable. There was every reason to believe, however, that the listeners were able discriminate among the different versions.

[17]In informal pilot work, the author has explored this issue with very short musical excerpts. It seems that, in judging the relative similarity of timing patterns across changes in tempo, a listener would not only be biased by the tempo difference as such but also would rely merely on local features (especially the initial and final IOIs) in making the judgment. (See also Handel, 1993.)

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 173-192

Detectability of Duration and Intensity Increments in Melody Tones: A Partial Connection between Music Perception and Performance*

Bruno H. Repp

Two experiments demonstrate positional variation in the relative detectability of, respectively, local temporal and dynamic perturbations in an isochronous and isodynamic sequence of melody tones, played on a computer-controlled piano. This variation may reflect listeners' expectations of expressive performance microstructure (the "top-down hypothesis"), or it may be due to psychoacoustic (pitch-related) stimulus factors (the "bottom-up hypothesis"). Percent correct scores for increments in tone duration correlated significantly with the average timing profile of pianists' expressive performances of the music, as predicted specifically by the top-down hypothesis. For intensity increments, the analogous perception-performance correlation was weak and the bottom-up factors of relative pitch height and/or direction of pitch change accounted for some of the perceptual variation. Subjects' musical training increased overall detection accuracy but did not affect the positional variation in accuracy scores in either experiment. These results are consistent with the top-down hypothesis for timing, but they favor the bottom-up hypothesis for dynamics. The perception-performance correlation for timing may also be viewed as being due to complex stimulus properties such as tonal motion and tension/relaxation that influence performers and listeners in similar ways.

Music played by human performers, Western tonal art music in particular, exhibits rich and finely differentiated variation that cannot be captured by conventional notation. This variation contributes vitally to the naturalness, expressiveness, and individuality of a performance. Collectively, it is known as *expressive microstructure* (Clynes, 1983). Its two most important dimensions are *agogics* and *dynamics*. The agogic (or timing) microstructure represents continuous modulations in local tempo or tone interonset intervals, whereas the dynamic (or intensity) microstructure represents the pattern of relative tone intensities (see Todd, 1992, 1995). The variation is not random but to a large extent rule-governed, despite much individual variability (see, e.g., Gabrielsson, 1987).

This research was supported by NIH Grant MH-51230. The author is grateful to Charles Nichols for his expert assistance, to Jonathan Berger for his permission to use the Yamaha Disclavier at the Yale University Center for Studies in Music Technology, and to Peter Desain, Henkjan Honing, Neil Macmillan, and particularly Mari Riess Jones for many helpful comments on the manuscript.

The primary purpose of performers' expressive devices is to elucidate the musical structure (Clarke, 1985; Palmer, 1989) and to create an allusion to physical or biological motion within this structural organization (Todd, 1992, 1995). Musically experienced listeners have corresponding expectations about how an expressive performance of a particular composition should be shaped. Experienced musicians' tacit knowledge of the rules governing expressive microstructure enables them to play expressively even when sightreading a new piece; experienced listeners' analogous knowledge enables them to appreciate and evaluate a performance, even of music not heard previously (as long as it is in a familiar style). The mental sound image of music imagined, remembered, or read from a score is almost certainly expressive, not mechanically rigid. Musical listeners' specific expectations about the expressive microstructure of a specific piece of music may account for the fact that expressive variation in well-performed music is usually not noticed as such; attention is drawn to the agogics or dynamics only when the variation is excessive

or goes in unexpected directions. Researchers working on computer synthesis of expressive performance have also observed this informally.

Based on these observations and considerations, Repp (1992a) devised an experimental method to assess listeners' specific microstructural expectations. He presented listeners with multivoiced excerpts from the piano literature, which were played with isochronous timing (i.e., with mechanically regular tone interonset intervals or IOIs) and *legato* (i.e., without any silent intervals between tones) on a computer-controlled digital piano. In each of the repeated presentations of an excerpt, one or two nonadjacent IOIs (as well as the tones filling them, to maintain *legato* articulation) were lengthened by a small amount, and the musically trained listeners' task was to detect and report the position of the lengthened tone(s). All IOIs in each musical excerpt were "probed" in this way, and the percentages of correct responses were plotted as a function of position to yield a *detection accuracy profile* (DAP) for each excerpt. A *false alarm profile* (FAP) based on incorrect responses was also derived. False alarm rates were thought to be a more direct (though less reliable) measure of subjects' expectations: An (unchanged) IOI expected to be relatively short should sound relatively long and hence attract false alarms. Finally, a representative *performance timing profile* was obtained from measurements of the expressive timing patterns of expert performances of the music. Repp's hypothesis was that the relative difficulty of detecting lengthening and the relative frequency of false alarms for each tone would both be inversely related to its relative degree of lengthening in a typical expressive performance.[1]

Correct response and false alarm rates indeed varied dramatically across positions and were positively correlated, indicating variable expectations or perceptual biases. Moreover, the predicted negative correlation between the DAP (and the FAP) and the performance timing profile was obtained: Lengthening was more difficult to detect in those positions where musicians were likely to slow down. Since performance timing was related to the musical structure, so was perception: Lengthening was observed in performance and was more difficult to detect (in an isochronous context) in metrically accented positions, close to the end of the excerpt, and at the ends of structural units (phrases and subphrases); moreover, both the extent of observed lengthening and the difficulty of detection increased with the depth of the nearest boundary in the hierarchical grouping structure (Lerdahl & Jackendoff, 1983).

These findings seemed to provide impressive evidence for the existence of microstructural expectations, at least with regard to timing (and lengthening in particular). However, the mechanisms by which these expectations reveal themselves in the laboratory remain a matter of speculation. It must be assumed that an isochronous musical excerpt, despite its deadpan quality and repeated presentation, automatically and instantly accesses a mental representation of which the expected microstructure is an integral part. Moreover, the expectations thus generated must interact immediately with veridical perception of timing, either directly by distorting the perceived durations of the IOIs or indirectly by affecting response decisions at some early, unconscious stage. These assumptions, although they are in the spirit of popular interactive processing models, are not without problems. For example, it is not clear why listeners do not establish a deadpan mental representation of the music after hearing it many times in the course of the experiment. In Repp's (1992a) study, there was no indication that the positional effects decreased over time. Also, according to his hypothesis—henceforth the "top-down hypothesis"—musically inexperienced subjects should not have well-defined microstructural expectations; yet, his experiments did not reveal a clear effect of musical experience.

A possible alternative account of his findings must therefore be considered—a "bottom-up hypothesis," according to which the variation in the DAP and FAP arises from psychoacoustic stimulus factors, without any reference to higher-level knowledge about musical structure and microstructure (cf. Monahan & Hirsh, 1990; Drake, 1993). Repp (1992a) made an attempt to assess the role of simple stimulus factors (pitch height and distance, absolute and relative intensity, tone density) in his materials via multiple regression analysis, but without any clear result. Yet, such bottom-up variables deserve further attention in view of recent findings that even young infants are sensitive to major phrase boundary cues in music (Krumhansl & Jusczyk, 1990; Jusczyk & Krumhansl, 1993).

The bottom-up and top-down hypotheses are not mutually exclusive, and they are difficult to separate conceptually and methodologically when the music is complex, because there are many

bottom-up cues to the higher-level structural representations that constrain observed microstructural variations in performance as well as listeners' expectations about these variations. In fact, it is arguable to what extent musical structure is in the sound pattern and to what extent it is a cognitive construct of performers and listeners (see the General Discussion).

The purpose of the present study was to re-examine the two hypotheses using simpler musical materials, where potential bottom-up accounts for variations in detection performance were more limited and could be defined more clearly. Instead of original piano compositions played with expressive dynamics (Repp, 1992a), the present experiments used simple monophonic tunes composed of piano tones of equal duration and intensity (except for tones that were detection targets), at the risk of attenuating the microstructural variations and expectations elicited by the materials and thus undermining the top-down hypothesis. In Experiment 1, the task was again the detection of duration increments, whereas Experiment 2 extended the investigation to the detection of intensity increments.

Because the only variable stimulus property (apart from the change to be detected) was pitch, specific bottom-up hypotheses were restricted to effects that pitch may have on perception of relative IOI duration or on the relative loudness of piano tones. In principle, such effects can take two forms: The pitch variation can result in variations in *sensitivity* across positions in the tune, such that changes in duration or intensity are more difficult to detect in some positions than in others, or it can introduce position-specific *perceptual bias*, such that some IOIs (tones) are perceived as a priori longer or louder than others. Both effects will affect the DAP, but only bias will affect the FAP as well. Thus the bottom-up hypothesis can account for a positive correlation between the DAP and the FAP (indicating variation in bias), but it is also compatible with the absence of such a correlation (indicating variation in sensitivity only).[2] The top-down hypothesis, on the other hand, necessarily implies a directional bias and hence is only compatible with a positive DAP-FAP correlation.

The major prediction of the top-down hypothesis is the negative correlation between the DAP and the performance profile. The bottom-up hypothesis does not predict such a correlation and has difficulty accounting for it. The correlation would have to be either coincidental or due to performers' attempts to compensate for perceptual biases introduced by bottom-up factors (Drake, 1993). At first blush, this seems implausible: Agogic and dynamic variation in expressive performance is generally much larger than seems necessary from this viewpoint, and performers do not generally have the goal of making their performance seem mechanically precise, as a compensation account would imply. However, it could be that bottom-up perceptual effects provide the seeds from which expressive strategies sprout as a form of deliberate exaggeration or overcompensation. While this suggestion is quite speculative, the bottom-up hypothesis deserves attention precisely because it could offer *explanations* for some expressive conventions. The top-down hypothesis, within the present experimental context, takes these conventions as given and inherent in cognitive structural representations of the music.

Specific bottom-up hypotheses for the present tasks can be derived from the existing psychoacoustic literature, despite considerable differences in stimuli and methodology. Psychoacoustic research characteristically uses extremely simple stimuli and highly practiced listeners. For example, studies of duration discrimination typically present silent intervals delimited by the onsets of very brief sounds, so that IOI and silent gap duration covary. In the present materials, however, long (600 ms), gradually decaying tones of varying pitch followed each other without intervening silence, and tone duration covaried with IOI. Moreover, the present listeners received no special training and were faced with high uncertainty about the location of the change to be detected.

An increase in the difficulty of duration discrimination with the pitch distance between two marker tones has been obtained in many studies using very short silent intervals (e.g., Perrott & Williams, 1971; Williams & Perrott, 1972; Collyer, 1974; Fitzgibbons, Pollatsek, & Thomas, 1974; Divenyi & Danner, 1977; Neff, Jesteadt, & Brown, 1982; Formby & Forrest, 1991). These intervals were an order of magnitude shorter than the filled IOIs in the present musical paradigm; also, short gaps are generally perceived as interruptions (i.e., as offset-onset intervals) rather than as onset-onset intervals. Moreover, Divenyi and Sachs (1978) found that the effect of pitch distance on the discrimination of silent intervals decreased with interval duration and was essentially absent at durations beyond 50 ms. Therefore, the relevance of these results to the present study is questionable.[3] However, there are

some indications that pitch distance effects also occur at longer IOIs.

Hirsh, Monahan, Grant, and Singh (1990: Exp. 2) presented their subjects with sequences of six 20-ms tones at IOIs of 200 ms and determined the just detectable delay in the onset of a single tone, which sometimes also differed in pitch from the other tones. The pitch difference tended to raise the discrimination threshold, but not consistently so; there were complex interactions with position in the sequence, and with the direction and magnitude of the pitch change. In a much earlier study with three-tone sequences, Divenyi (1971) already observed that the detectability of timing perturbations was not a simple function of the frequency separation between tones. In particular, he found that lengthening of the silence between tones was more difficult to detect when the frequencies formed a simple ratio, i.e., a common musical interval. These effects, however, again tended to wash out at slower rates of presentation.

In a recent experiment similar in motivation to the present study, Drake (1993) presented listeners with simple melodic sequences composed of six 50-ms pure tones at IOIs of 300 ms. The sequences contained either two pitch jumps (C-C-G-G-C-C) or a pitch turn (D-E-F-E-D-C). The (untrained) subjects' task was to detect and locate changes in duration (both increments and decrements) in any of the five IOIs. Changes at pitch jump locations were more difficult to locate (but not more difficult to detect) than changes in other positions; there was no effect in the pitch turn sequence. Thus there is no very clear evidence so far that pitch distance has an effect on *sensitivity* to temporal change at relatively long IOIs.

More convincing evidence that pitch distance can create a *perceptual bias* comes from studies of the auditory "kappa effect" (Shigeno, 1986, 1993; Crowder & Neath, 1995). In this paradigm, listeners are asked to compare the durations of two time intervals delimited by three tones of different frequency. The consistent finding is that, when the frequency of the second tone is closer to that of the first tone than to that of the third tone and the two time intervals are equal, subjects perceive the first time interval to be shorter than the second. The interval durations in these tasks were comparable to those employed in the present study, but the tones were separated by silence rather than contiguous. It is not clear whether the kappa effect applies to the IOIs or to the silences between tones.

Despite these difficulties of generalization, there seems to be only one reasonable bottom-up hypothesis for the present duration increment detection task: An increment in IOI duration may be more difficult to detect if the tones delimiting it are widely separated in pitch than if they are close in pitch. This could be due either to reduced sensitivity or to a bias (viz., the kappa effect), with different consequences for the DAP-FAP correlation. There should be a negative correlation between the DAP and the absolute pitch distances (i.e., regardless of direction) between successive tones in the tune.

As to the possible effect of pitch distance on intensity discrimination, there is surprisingly little relevant psychoacoustic literature. Nearly all intensity discrimination tasks have used carriers with identical spectral characteristics. Dai and Green (1992) have demonstrated that intensity differences between two successive pure tones are more difficult to detect when the tones fall into different critical bands. However, it is not known whether this finding would generalize to complex tones differing in fundamental frequency, whose spectra overlap extensively.[4] Drake (1993) included an intensity discrimination task in her study (referred to above) and found poorer detection of intensity increments, but better detection of intensity decrements, on the high notes in her pitch jump sequence. This suggests a perceptual bias to perceive higher tones as less loud than lower tones, but again the generalizability of these results to complex musical sounds is not guaranteed. It is noteworthy, however, that the generative rules for music performance developed by Sundberg and his collaborators include a 3 dB/octave increase in sound level with pitch (see Sundberg, 1988; Friberg, 1991). This rule could represent either a deliberate effort to compensate for a reduced perceived loudness of higher complex tones (a bottom-up effect), or an attempt to satisfy listeners' expectations about typical performance dynamics (a top-down effect). Such expectations could derive from a correlation of pitch and dynamics in music performance.

It may be hypothesized, then, that intensity increments will be more difficult to detect in a high tone than in a low tone, and perhaps also that detection scores will be in inverse proportion to the pitch distance from the preceding tone. These bottom-up hypotheses predict negative correlations between the DAP and the absolute

pitches of, or the absolute pitch distances between, successive tones in the tune. While pitch distance may affect sensitivity, absolute pitch may cause a bias to hear lower tones as louder. However, if there is also a tendency to play higher tones louder in performance, the resulting negative DAP-performance correlation would be compatible with either a bottom-up or a top-down account.

Additional, local bottom-up effects predicted for each task are that detection of a change in the first and last IOI or tone of a melodic sequence should be impaired, due to the absence of one adjacent IOI or tone for comparison. In duration increment detection, the final IOI should suffer especially (Hirsh et al., 1990; Monahan & Hirsh, 1990). In intensity discrimination, the initial tone may be more affected. (Moreover, the final long tone was not "probed," as explained below.) A gradual increase in detection scores over the first four or five positions may be predicted on the basis of increasing perceptual definition of the standard IOI duration (Drake & Botte, 1993; Ivry & Hazeltine, 1995).

The melodies used included two features that were intended to provide additional fuel for the bottom-up and top-down accounts, respectively. As will be seen shortly, each melody was composed of three similar parts, each reaching an apex at a successively higher pitch and with a larger upward jump to that pitch. This systematic variation of pitch height represented a bottom-up factor that might affect duration and/or intensity increment detection. The two melodies also had almost identical pitches but had different metrical properties, induced by differences in the notation and in the exact sequence of pitches. Metrical structure (by which is meant here the placement of the theoretical downbeats in the pitch sequence) was a pure top-down variable since no temporal or dynamic accents were present in the stimuli; therefore, any effect of metrical structure on the DAP and FAP was going to be an additional indicator of top-down expectations, provided that metrical structure also affected performance. If performance microstructure happened to be unaffected by metrical structure, then, according to the top-down hypothesis, listeners should be insensitive to it also. Effects of metrical structure on performance but not on perception, or vice versa, would be inconsistent with the top-down hypothesis. The bottom-up hypothesis, of course, does not predict any effects of meter.

Finally, an important factor relevant to both hypotheses was re-examined in the present study, namely the role of listeners' musical experience. According to the top-down hypothesis, musically experienced listeners should have more clearly defined expectations about performance microstructure and thus should show a more finely differentiated DAP that is more highly correlated with the relevant performance profile than the DAP of musically untrained subjects. The hypothesis makes no predictions about absolute accuracy, as strong expectations may actually hinder detection. The bottom-up hypothesis, on the other hand, predicts only higher accuracy for musically experienced subjects, because of their training or superior auditory abilities, but no difference in the DAP. As mentioned earlier, Repp (1992a) did not find any very clear effects of musical training on either overall accuracy or on the DAP, despite a wide range of accuracy scores. In precursors to the present experiments, Repp (1992b) found a correlation between musical experience and the overall intensity increment detection score, but no effect on the DAP in either duration or intensity increment detection, which was problematic for the top-down hypothesis. These previous studies used heterogeneous groups of subjects and examined correlations with questionnaire measures of musical experience. Here, a more focused approach was taken by sorting subjects into groups according to musical training and comparing them by means of ANOVA.

PERFORMANCE MEASUREMENTS

Musical materials

The two experimental tunes (which may also be considered as two versions of a single tune), labelled Tune A and Tune B, are shown in Figure 1 in musical notation. They were composed by the author with the intention of providing melodies that invited expressive performance. Both began with a *staccato* note which served to mark the first downbeat in the perceptual experiment. This initial note was followed by a quarter-note rest and a double upbeat in Tune A, but by two quarter-note rests and a single upbeat in Tune B. By defining the duration of the rest, the initial tone thus served as a prime for the metrical structure of the tunes. All subsequent notes were quarter notes, except for the final long note.

The pitches in the two tunes were almost identical; they described three cycles of an up-down pitch motion, corresponding to three subphrases. As can be seen, however, the turning points in the pitch contour were shifted by one beat in Tune B relative to Tune A: In Tune A, they always occurred on a downbeat; in Tune B, they always occurred on the metrically weak second beat.

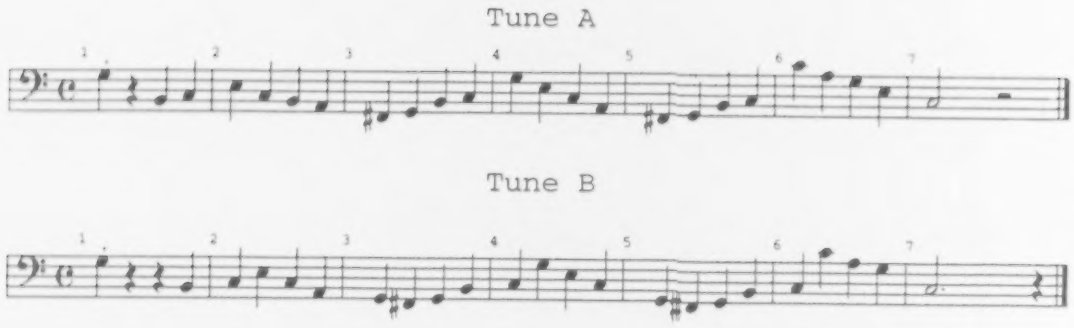



Figure 1. Musical materials for both experiments. The small digits are bar numbers.

Similarly, the subphrase boundaries occurred at different metrical points in the two tunes: following positions 3-2 and 5-2 in Tune A, but following positions 3-3 and 5-3 in Tune B.[5] However, the subphrase boundaries occurred at the same points within the pitch structure of each tune. In other words, the grouping structure was aligned with the pitch structure, but the metrical structure was shifted with respect to these two.

The three subphrases differed in the height of the upper turning point, which represents an upward excursion of a third in the first subphrase, of a fifth in the second subphrase, and of an octave in the third subphrase. The following note, however, was always a (major or minor) third lower, so the downstep from the pitch peak was essentially held constant. The focus thus was on upward pitch jumps.

The small differences in pitch structure between Tunes A and B, together with the notation and the initial priming tone, served to force listeners into a particular metrical framework. If the pitches had been identical in the two tunes, their metrical structure would have been ambiguous. In the present materials, although a listener could start out hearing one tune with the metrical structure of the other, this interpretation would lead to what are arguably less well-formed melodies, and at the end there would either be an extra note or a missing note. It was expected, therefore, that such an awkward metrical interpretation would be abandoned after a few hearings.

Method

Five pianists performed the experimental tunes. Four of them were graduate students of piano performance at the Yale School of Music, and the fifth was the author, a serious amateur. After a short practice period, each pianist played the two tunes from the notation (Figure 1) three times in alternation, with the right hand, at a moderate tempo, *legato*, and "with expression." The expressive shaping was done intuitively, without conscious deliberation of the musical structure or microstructure. The different metrical structure of the tunes was obvious from the notation and was not pointed out specifically.

The graduate students played on a Yamaha MX100A Disclavier, which is a real (i.e., mechanical-acoustic) upright piano with added electronic components that enable computer recording and playback of performances. The author played on a Roland RD-250s digital piano with "Piano 1" sound, monitored over earphones. The onset times and velocities of all keystrokes (as well as their offset times, which are irrelevant here) were registered by a microcomputer in MIDI format. Interonset intervals (IOIs) were calculated from the onset times. Velocities were represented by numbers between 0 and 127 which, in the relevant mid-range, correspond to steps of about 0.25 dB in peak rms sound level (Repp, 1993a).

Results and discussion

The IOIs and velocities were averaged, first over the three repetitions and then across the five pianists' performances of each tune.[6] Figure 2 shows these average timing and intensity profiles. In view of the extensive pitch commonality of the two tunes, their profiles have been superimposed in each panel of the figure. The subphrase boundaries are indicated by vertical dotted lines. The abscissa shows the sequence of musical pitches. Pitches that occur in only one tune are represented by gaps in the other tune's profile. Of course, these pitches had to be omitted from statistical analyses comparing the two profiles.

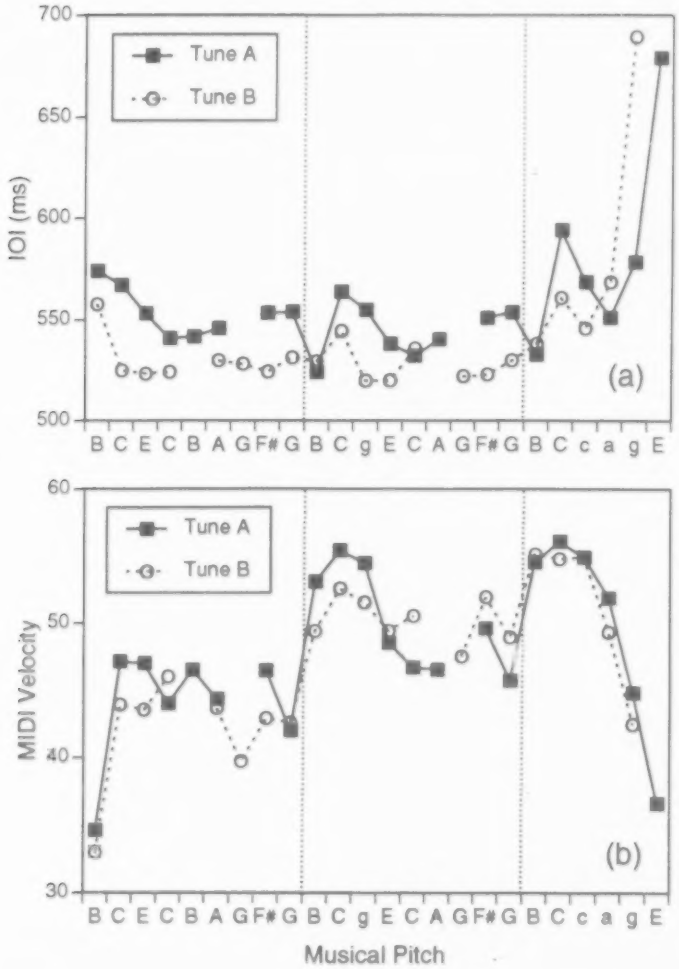


Figure 2. **Average IOIs and MIDI velocities of five pianists' performances of the two tunes used in Experiments 1 and 2, aligned according to musical pitch (assuming the key of C major). Vertical dotted lines indicate subphrase boundaries. Lower case letters indicate "an octave higher," not a particular octave.**

The average timing profiles (Figure 2a) were fairly flat, except for a pronounced *ritardando* at the end. Presumably, this was one price to be paid for using such simple materials. The most striking difference between the timing profiles for the two tunes occurred on the last shared pitch (g), which occupied the penultimate IOI in Tune A but the final IOI in Tune B. This difference thus occurred because the final *ritardando* was differentially aligned with the common pitches, due to the extra note (E) in Tune A. The time course of the *ritardando* was in fact quite similar in the two tunes. Therefore, the last shared note (g) was omitted from the ANOVA.

A two-way repeated-measures ANOVA on the IOIs (fixed factors: tune, position; random factor: pianists) revealed significant effects of position [$F(17,68) = 3.30$, $p < .0003$] and of tune [$F(1,4) = 33.46$, $p < .005$], as well as a position by tune interaction [$F(17,68) = 3.43$, $p < .0003$]. The position main effect indicates that there was reliable timing variation, apart from the final *ritardando*. The tune main effect indicates that Tune B tended to be played faster than Tune A, which is of little interest. The interaction, which is the effect of prime interest, seems to be mainly due to a shortening of the IOIs immediately following the subphrase boundaries in Tune A, but not in Tune B. In Tune A, the notes corresponding to these IOIs (pitch B) initiated double upbeats on the third beat, whereas in Tune B they formed single upbeats on the fourth beat (see Figure 1).

Their shortening in Tune A suggests that they were treated as upbeats to upbeats, as it were. A tendency present in both tunes is the progressive lengthening of the tone preceding the upward pitch jump in each subphrase (pitch C), as if it took longer to reach a higher pitch. There was also a slight acceleration at the beginning of each tune, which seemed to last longer in Tune A than in Tune B. Somewhat unexpectedly, there was no noticeable *ritardando* preceding subphrase boundaries. Thus the effects of metrical structure on performance timing were quite limited, being restricted to the onsets of subphrases.

The relationship between IOI duration and the absolute pitch distance between the tones defining the IOI was also examined. These correlations were positive but fell short of significance (0.37 and 0.40 for Tunes A and B respectively). Omission of the extended final IOI made the correlation significant in Tune A (0.60, $p < .01$) but not in Tune B (0.34). Thus there was a weak tendency to lengthen the IOI between tones far apart in pitch.

It appears that the pianists made greater use of dynamics than of agogics in performing the experimental melodies. The intensity profiles (Figure 2b) show striking variation across positions [$F(18,72) = 5.40$, $p < .0001$], but there was no significant tune by position interaction [$F(18,72) = 1.00$] and hence no effect of metrical structure.[7] The profiles show a pronounced peak on the tones involved in the upward pitch step at the beginning of each subphrase, and this peak increases in height from the first to the second subphrase, but little thereafter. Correlations between absolute pitch height and dynamics were positive but small. Somewhat larger, but still nonsignificant positive correlations were obtained with the absolute pitch distance from the preceding tone. The measure that correlated most strongly with dynamics was the *directional* pitch distance from the preceding tone (Tune A: 0.57, $p < .01$; Tune B: 0.43, $p < .05$). Thus there was a tendency to play louder when the pitch went up than when it went down.

To the extent that these performance profiles are representative, they provide an estimate of the expectations that the top-down hypothesis attributes to musically experienced listeners. The virtual absence of effects of metrical structure on performance was surprising, given that the pianists (including the author!) were well aware of the difference and played the tunes in alternation, which should have encouraged contrasting interpretations. Sloboda (1983, 1985) found that pianists could convey metrical structure through performance parameters, but more so through dynamics than through timing. However, his bouncy melodies were of a very different character than the present "expressive" tunes, which moved much more slowly, at a beat rate close to the optimal pulse (Fraisse, 1982; Parncutt, 1994), with only one tone per beat. This slow event rate and the resulting absence of a hierarchical rhythmic structure may have been responsible for the near-absence of metrical effects in the present case. It was predicted, therefore, that meter would have little effect in perception also. Even without metrical effects, there was enough variation in the performance profiles, especially in the intensity profile, to permit a fair assessment of the top-down hypothesis.

EXPERIMENT 1

Method

Subjects. Twenty-four paid volunteers participated in the study. They were divided into three groups of eight: musicians (M), amateur musicians (A), and nonmusicians (N). Subjects in group M had had at least eight years of formal training on an instrument and still played that instrument.[8] They included four graduate students at the Yale School of Music and four Yale undergraduates. They represented various instruments and ranged in age from 18 to 25. Subjects in group A had had some formal musical training (2-10 years) and were able to read music, but most of them did not play an instrument any more. They were mostly Yale undergraduates and ranged in age from 18 to 28 years, except for one subject who was 43. Subjects in group N had had little or no musical training, and only one of them could read music. They ranged in age from 19 to 33 years and were Yale undergraduates or employees.[9]

Materials. The tunes (Fig. 1) were realized on a computer-controlled Roland RD-250s digital piano with "Piano 1" sound, *legato* articulation, a standard IOI duration of 600 ms, and a constant MIDI velocity. The initial *staccato* tone only helped to establish the meter; the relevant IOIs were those between the subsequent tones. Completely isochronous versions were used for initial familiarization only. Experimental stimuli contained one or two IOIs that were lengthened by delaying the nominal offset of the tone occupying it and the onsets of all following tones in the MIDI instructions. The IOI following a target IOI thus remained unchanged. The lengthening of the tone

filling the target IOI was necessary to maintain *legato* articulation; otherwise, there would have been a salient alternative cue in the detection task. There were 22 target IOIs in Tune A and 21 in Tune B; these were "probed" in 15 and 14 trials, respectively. Thus about half the trials had two target IOIs which always occurred in different subphrases and were never very close to each other; also, they always occurred in different metrical positions. The assignment of target IOIs to trials was random. The duration increment to be detected was 8.3% (50 ms), 6.7% (40 ms), 5% (30 ms), and 3.3% (20 ms) during four consecutive test blocks, each containing 15 or 14 trials. The four test blocks of progressive difficulty were preceded by three completely isochronous examples of the tune and three demonstration trials with 10% (60 ms) lengthening. There were separate tests for Tune A and Tune B, which were identical except for the difference in number of trials. Trials were separated by a silent interval of about 4 seconds. To create variety, the tune was randomly transposed from one trial to the next within an octave range centered on the pitches shown in Figure 1. The test sequences were recorded directly from the digital piano onto digital tape.

Procedure. Subjects were tested individually in a quiet room. They listened over Sennheiser HD 540 II earphones and entered their responses on answer sheets that showed the tune in musical notation for each trial. The notated tune was always in C major, but the random transposition of the test stimuli was pointed out to the subjects. Subjects who could not read music were told that note height represented pitch height, and were asked to follow the score with their pencil as they listened. For the three demonstration trials, the correct responses had already been filled in. If subjects had difficulty hearing the lengthened tones, they were allowed to listen to these trials again. For the subsequent test blocks, the subjects were informed that there could be either one or two lengthened tones on each trial (never the first or last tone), and were asked to circle the note(s) corresponding to the lengthened tone(s). They were specifically asked not to circle the following note (a very common occurrence in the earlier studies of Repp, 1992a, 1992b) and not to guess randomly but to place a question mark at the end of the line if they could not hear any lengthened tone. Half the subjects in each group listened to the Tune A test before the Tune B test; the others listened in the reverse order. Before the second test, the different metrical structure of the new tune was explained carefully. Each test took about 22 minutes, and there was a break in between.

Results and discussion

Overall accuracy. Despite the explicit instructions, subjects again had a strong tendency to circle the note following the correct one ("late responses"), though individual differences were very large in that respect and no subject gave late responses exclusively . Therefore, responses were accepted as correct if they were adjacent to the correct position. Overall, among the 49.6% responses scored as correct, there were 32.8% "direct hits," 14.5% late responses, and only 2.3% "early responses."

There was a difference between the musicians and the other subjects in the incidence of late responses: Late responses were less than half as frequent in group M than in groups A and N, and since musicians gave more correct responses, this difference was even larger in terms of the average percentage of correct responses that was late: 14.8% (range: 1% to 48%) in group M versus 36.9% (range: 8% to 89%) in group A and 36.8% (range: 1% to 79%) in group N. Because of the enormous individual variability, the group difference did not reach significance in an ANOVA. However, 5 of the 8 subjects with rates below 10% were musicians, whereas none of the 6 subjects with rates above 50% was a musician. Two factors may underlie the late response tendency: (1) Listeners obviously hear the following tone when they realize that a lengthening has occurred, and they may circle the tone they hear instead of backtracking on the answer sheet. (2) They may attribute the perceived hesitation to the following tone because of its delayed onset, or possibly because the delayed onset makes it seem slightly accented (cf. Clarke, 1985).[10] Recent experiments (Repp, 1995c) suggest that both factors play a role.

The overall percentages of correct responses, averaged across positions and tunes, are shown in Figure 3 as a function of duration increment (test block), separately for the three groups of subjects. A repeated-measures ANOVA was carried out with the fixed factors group, block, tune, and order; subjects nested within groups constituted the random factor. Not surprisingly, performance declined significantly across test blocks [$F(3,18) = 168.49$, $p < .0001$]. There was also a significant main effect of subject group [$F(2,18) = 4.16$, $p < .04$]: Musicians performed better than amateurs and nonmusicians, with little difference between the latter two groups.[11] Since the guessing probability in this task is very small, it is clear

that subjects can detect 3.3% increments with better than chance accuracy.[12]

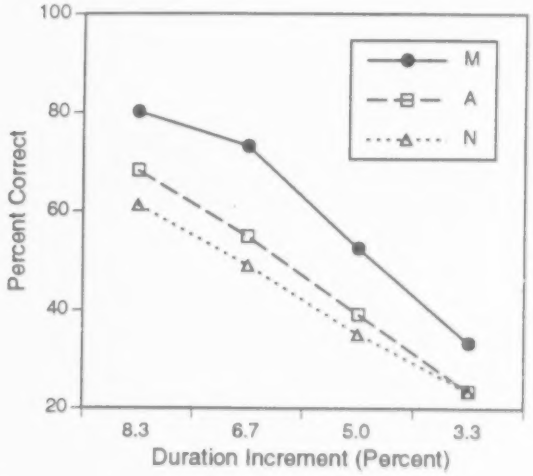


Figure 3. **Percent correct scores for the three subject groups in Experiment 1, averaged across positions, as a function of duration increment (test block).**

Detection accuracy profiles. Figure 4 shows the DAPs for Tunes A and B, separately for the three groups of subjects. Separate repeated-measures ANOVAs were conducted on each tune, with the fixed factors of group and position and the random factor of subjects nested within groups. Detection scores varied significantly as a function of position in both Tune A [$F(21,441) = 11.71$, $p < .0001$] and in Tune B [$F(20,420) = 11.32$, $p < .0001$]. There was a significant main effect of subject group for each tune, but the group by position interactions were far from significance. Figure 4 shows that, contrary to the prediction of the top-down hypothesis, the average DAP of the musicians was not more differentiated than that of the nonmusicians—if anything, it was a little flatter.

Figure 5 compares directly the average DAPs for Tunes A and B. The data are collapsed here over all subjects, as there were no significant interactions involving subject groups. The two tunes are aligned here by pitch normalized to the key of C, as in Figure 2. Any pitches that occur in one tune but not in the other are visible as gaps in the graph; they were omitted from the ANOVA reported below. The common subphrase boundaries are indicated by vertical dotted lines.

The similarities between the two profiles are more striking than the differences: Accuracy was low at the beginning, highest during the second subphrase, and declined during the final subphrase. The final decline replicates earlier findings (Repp, 1992a, 1992b) and may be attributed to listeners' expectation of a final *ritardando*, in agreement with the top-down hypothesis.

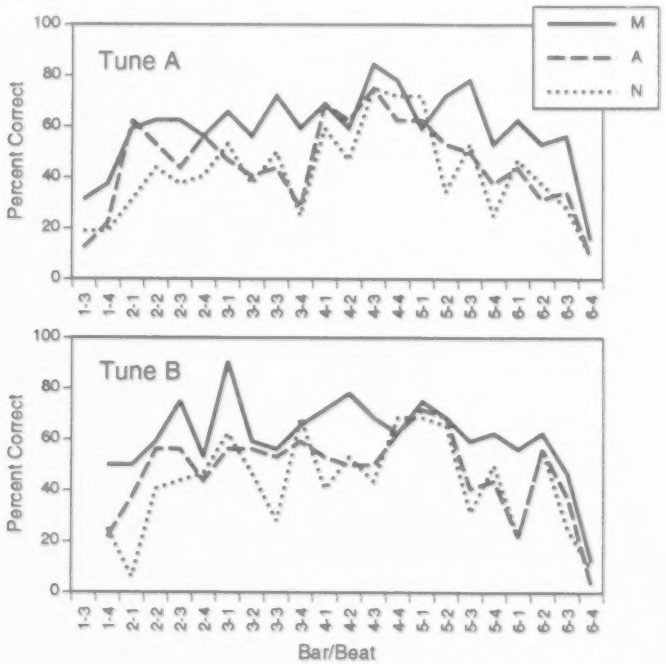


Figure 4. **Detection accuracy profiles for Tunes A and B in Experiment 1, separately for each of the three subject groups.**

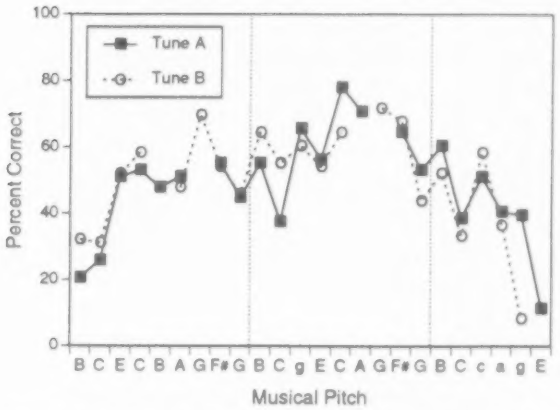


Figure 5. **Average DAPs for Tunes A and B in Experiment 1, aligned according to pitch. Vertical dotted lines indicate subphrase boundaries.**

The bottom-up hypothesis predicts only the poor performance in the final position, but not the preceding decline. The poor performance in the initial two positions is consistent with a bottom-up explanation, though a more gradual increase in accuracy was expected. There are small dips in the profiles at analogous points in the second and third subphrase, immediately preceding the upward pitch jumps (pitch C); the low performance in the second position may also reflect this bottom-up factor. The magnitude of the jump, however, did not seem to matter. Note also that the direction of the effect is not consistent with the kappa effect, according to which IOIs at pitch jump locations should sound longer to begin with, thus making duration increments easier to detect.

The most obvious difference between the two DAPs is at the last common pitch, g, where detection scores were 31% lower in Tune B than in Tune A. This tone fills the final IOI in Tune B, but it is followed by another IOI in Tune A. For that IOI, performance was as low as for the final IOI in Tune B. Note the corresponding (inverse) difference in the average performance timing profiles (Fig. 2a) and its discussion above. As in the performance analysis, the last common pitch was omitted from the ANOVA.

A repeated-measures ANOVA (fixed factors: group, tune, position; random factor: subjects nested within groups) yielded, in addition to the expected main effects of position and subject group, a marginally significant tune by position interaction [$F(17,357) = 1.72$, $p < .04$]. The largest remaining difference between the profiles occurred in the second IOI of the second subphrase (pitch C), where performance was 18% lower in Tune A than in Tune B. This looks like an effect of metrical structure (the tone falls on a downbeat in Tune B but on the weak fourth beat in Tune A), and it also mirrors an effect noted in the performance timing profile (Fig. 2a); however, the analogous positions in the first and third subphrases show no such difference. Thus, there are no consistent effects of metrical structure. Since metrical effects were also rather weak in performance, this finding does not upset the top-down hypothesis.[13]

As predicted by the top-down hypothesis, the average DAPs (Fig. 5) were significantly correlated with the average performance timing profiles (Fig. 2a). The correlation was -0.74 ($p < .001$) for each tune. It seems that the last data point must have contributed substantially to these correlations. With that data point omitted, however, the correlations were still significant: -0.66 ($p < .001$) for Tune A and -0.57 ($p < .01$) for Tune B. A local inverse correspondence may also be seen in the performance and accuracy profiles for Tune A at the beginnings of the second and third subphrases. However, while the performance data suggested a possible effect of metrical structure at this point because Tune B showed less of a timing perturbation than Tune A, the DAPs show similar results for the two tunes. Thus the perceptual effect is probably not metrical in nature; it may reflect expectation of lengthening preceding an upward pitch jump. The height of the jump seemed to matter little, however.

The top-down hypothesis also predicted that musicians would show stronger perception-performance correlations than nonmusicians. A tendency in that direction was in fact obtained, but only when the final data point was included. In that case, the correlations for the three subject groups were -0.79, -0.66, and -0.64 for Tune A, and -0.80, -0.76, and -0.52 for Tune B. With the final data point omitted, however, this tendency disappeared. Thus it cannot be given much weight.

The main prediction of the bottom-up hypothesis was that the relative detectability of lengthening would be inversely related to the pitch distance between the tones delimiting an IOI. The correlations of the DAPs with this variable (expressed in semitones) were indeed negative but significant only for one tune (Tune A: -0.27, n.s.; Tune B: -0.62, $p < .01$). This gives only weak support to the hypothesis. Moreover, since a weak positive correlation between absolute pitch distance and IOI duration was observed in performance, the present negative correlation is compatible with a top-down as well as a bottom-up account.

False alarm profiles. Because of the possibility of two target IOIs on a trial, subjects were free to

give two responses on each trial, if they wanted. However, false alarms (circling of any note not adjacent to a correct position) were relatively infrequent: Relative to the total number of target positions, only 8% false alarm responses were given; relative to the number of missed targets (50.4%), the false alarm rate was 16%.[14] Individual differences were considerable, ranging from 0.6% to one exceptional case of 36.9% false alarms relative to the number of target positions. There were no pronounced differences among the three subject groups, but for some reason Tune A elicited consistently more false alarms (8.9%) than Tune B (7.2%) [$F(1,21) = 5.83, p < .03$].

The FAPs are shown in Figure 6. Note that this figure plots numbers of responses, not percentages. The profiles for the two aligned tunes were quite similar, except for the very high false alarm rate on the first g in the second subphrase in Tune A. It is possible that the coincidence of that tone with a downbeat enhanced the false alarm tendency, but there are no such meter-related differences in other corresponding locations. It appears that the high-pitched notes following pitch skips attracted false alarms, but the octave jump in the third subphrase caused fewer false alarms than the jump of a fifth in the second subphrase. Apart from these locations, false alarms were also frequent on pitches A and G in the second subphrase, where G fell on a downbeat in Tune B, but A fell on the weak fourth beat in Tune A. Thus, again, there is no clear evidence for any effect of metrical structure. Because of the large individual differences in false alarm rates, no ANOVA was conducted on the FAPs.

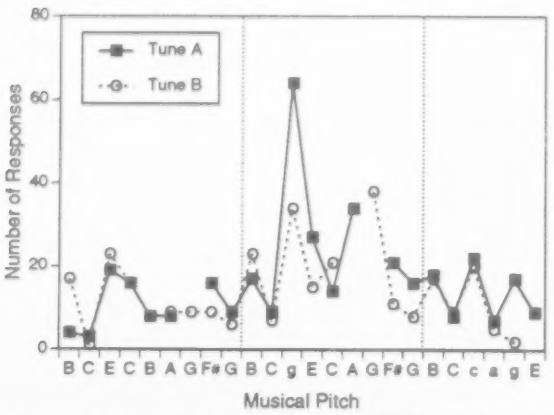


Figure 6. **Average false alarm profiles for Tunes A and B in Experiment 1, aligned according to pitch.**

The peaks and valleys in the FAPs match those in the DAPs (Figure 5) rather well, though their relative height differs. The correlations were significant, 0.57 for Tune A and 0.60 for Tune B, both $p < .01$. These correlations are consistent with the top-down hypothesis: False alarms tend to occur on those IOIs that, because they are not expected to be lengthened, sound relatively long to begin with. The correlations would also be consistent with the bottom-up hypothesis if false alarm rates showed a positive correlation with the absolute pitch distance between tones delimiting the IOIs (i.e., the kappa effect). However, these correlations were *negative* and nonsignificant (Tune A: -0.11; Tune B: -0.31). Thus, if anything, subjects perceived IOIs as longer when they were between tones close in pitch. This finding suggests that the kappa effect did not operate in the present tunes, and hence it provides no support for the bottom-up hypothesis.

Summary. The results of Experiment 1, obtained with much simpler materials than used by Repp (1992a), provide additional support for the main prediction of his top-down hypothesis: Detectability of lengthening was negatively correlated with lengthening in performance, and false alarm frequencies (a more direct but less precise index of listeners' expectations) correlated positively with detection accuracy. Subjects with extensive musical training performed better overall but did not have more differentiated DAPs, contrary to a secondary prediction of the hypothesis. The finding that effects of metrical structure were generally absent in both perception and performance, while disappointing, is not inconsistent with the top-down hypothesis. The bottom-up hypothesis is consistent with these secondary results but does not provide a convincing account of the main findings. In particular, it provides no explanation of the gradual decline in detection performance towards the end of a tune, and it provides no rationale for the false alarm distribution because the kappa effect seems to be absent.

EXPERIMENT 2

This experiment was quite analogous to Experiment 1, except that intensity increments rather than duration increments were to be detected. The very distinctive performance intensity profile (Figure 2b) provided a good basis for reexamining the top-down hypothesis for intensity increment detection, which had not received much support in a pilot study (Repp, 1992b).

Method

Subjects. Twenty new subjects were recruited from the Yale community and divided into three groups according to the same criteria as in Experiment 1. There were 8 musicians (aged 20-

33), 6 amateurs (aged 21-38), and 6 nonmusicians (aged 29-45). The differences in mean age among the groups [$F(2,17) = 4.52$, $p < .03$] were inadvertent (volunteers were taken as they came) but were not considered a serious problem.[15] The subjects were paid for their services.

Materials. The tunes and the test arrangement were the same as in Experiment 1; even the same random sequences were used. The only difference was that intensity increments occurred instead of duration increments.[16] The initial *staccato* tone and the final long tone were not possible targets. The increment was 11 MIDI velocity units (about 2.75 dB) in the three demonstration trials, and 9 (2.25 dB), 7 (1.75 dB), 5 (1.25 dB), and 3 velocity units (0.75 dB) in the four test blocks. Tunes were again randomly transposed from trial to trial, which was especially important in this experiment, as there was some random variability in peak sound level among digital piano tones of different pitch (Repp, 1993a).

Procedure. The procedure was the same as in Experiment 1, except that the subjects were asked to circle the notes corresponding to tones that seemed louder than the others. The possibility of two targets on a trial was pointed out. Instructions not to circle the following note were omitted, as no such tendency was expected in this task. The order of the two tunes was nearly counterbalanced: In the nonmusician group, four subjects happened to listen in one order and two in the other.

Results and Discussion

Overall accuracy. As in the previous experiments in this series, a liberal scoring criterion was adopted accepting responses to adjacent positions as correct, even though they were quite rare: Of the 45.7% correct responses, 2.6% were early and 1.9% late. The decision to count these responses as correct seems justified, for although they probably included some random guesses (as did direct hits, of course), individual and group differences in their frequency suggested that they were at least partially made up of misplaced correct responses. Individual percentages of such responses ranged from 0 to 10.5. They were twice as frequent in groups A and N than in group M, though that group difference was not significant. The difference in frequency of early and late responses was not significant either. Surprisingly, however, there was a main effect of tune and a tune by group interaction [$F(2,17) = 7.27$, $p < .006$]. These effects were due to the fact that nonmusicians were much more likely to misplace their responses in Tune B than in Tune A, for unknown reasons.

The overall percentages of correct responses are shown in Figure 7. Not surprisingly, detection performance declined as the increment got smaller [$F(3,42) = 92.82$, $p < .0001$]. However, even in the most difficult condition scores were still above chance, certainly for the musicians.[17] As in Experiment 1, there was a significant difference between subject groups, with musicians performing best and nonmusicians worst [$F(2,14) = 7.58$, $p < .006$]. Because of the unintended confounding of musical experience with age, age was entered as a covariate in a follow-up analysis. The group difference remained significant [$F(2,14) = 6.56$, $p < .009$], and the correlation between age and accuracy scores was nonsignificant. (Among the nonmusicians, the oldest subject had the highest score.)

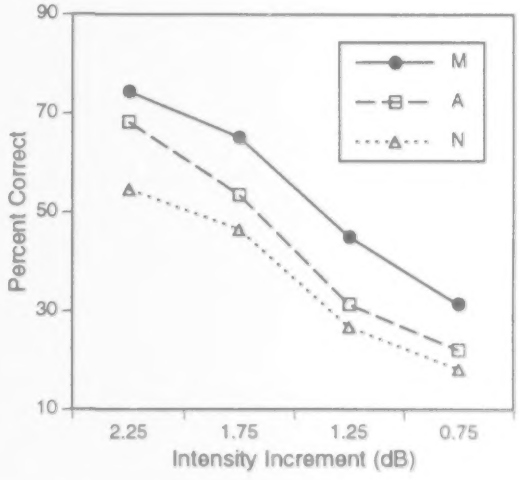


Figure 7. **Percent correct scores for the three subject groups in Experiment 2, averaged across positions, as a function of intensity increment (test block).**

Detection accuracy profiles. The DAPs of the three subject groups are shown separately for each tune in Figure 8. It can be seen that there was dramatic variation in accuracy across positions; the position main effect was highly significant for both Tune A [$F(21,357) = 14.12$, $p < .0001$] and Tune B [$F(20,340) = 12.34$, $p < .0001$]. In addition, each tune showed a group main effect [Tune A: $F(2,17) = 4.39$, $p < .03$; Tune B: $F(2,17) = 8.92$, $p < .003$]. However, even though Figure 8 seems to suggest that group differences were larger during the second half of each tune than during the first half, the position by group interaction was far from significance in each case. Again, there is no evidence that musicians showed a more differentiated DAP than amateurs or nonmusicians.

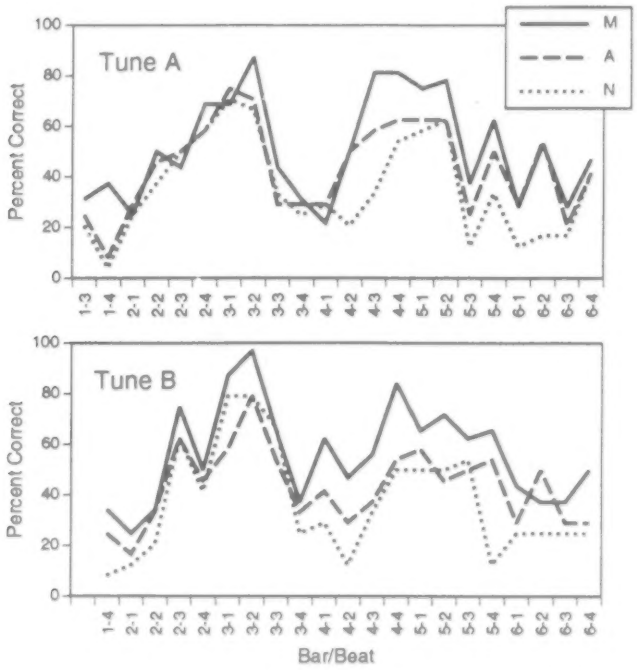


Figure 8. Detection accuracy profiles for Tunes A and B in Experiment 2, separately for each of the three subject groups.

The average DAPs of the two tunes are shown aligned according to pitch in Figure 9. As in Experiment 1, the two profiles were very similar. The subphrase boundaries are indicated by vertical dotted lines. In each tune, intensity increments were difficult to detect at the beginning of a subphrase and much easier during its second half (which was missing in the third, abbreviated subphrase).

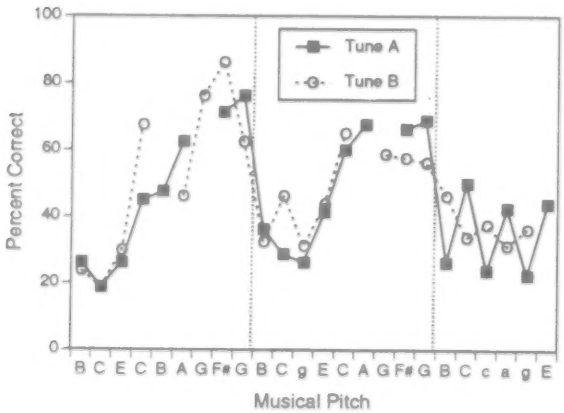


Figure 9. Average DAPs for Tunes A and B in Experiment 2, aligned according to pitch.

The ANOVA on the aligned profiles naturally showed a highly significant effect of position [$F(18,306) = 30.03$, $p < .0001$] as well as a main effect of subject group [$F(2,17) = 10.37$, $p < .002$], with no main effect of tune. However, the tune by position interaction was significant [$F(18,306) = 3.67$, $p < .0001$], indicating that the two functions, while similar, were not identical. Could these differences have been due to effects of metrical structure? Even though such effects were not evident in performance, a possible prediction was that intensity increments should be more difficult to detect on downbeats or in strong metrical positions in general (i.e., the first and third beats in each bar). Tune A shows a zigzag pattern during the last subphrase that is consistent with this prediction, as the valleys in the DAP coincide with strong metrical positions. However, Tune B does not show the complementary pattern. Moreover, there is no evidence of metrical effects elsewhere in the tunes. Consider, for example, the adjacent pitches C and E in the first subphrase (second and third data points): E fell on the downbeat in Tune A, C on the downbeat in Tune B, yet there is no interaction. The analogous pair of pitches in the second subphrase, C and g, shows a difference on C in favor of Tune B; that

difference is in the wrong direction, however, as C was on the downbeat in Tune B and should have shown lower performance than in Tune A. Only the third analogous pair in the last subphrase, C and c, shows the predicted interaction. Overall, however, there is no convincing evidence for effects of metrical structure, which—in view of the absence of metrical effects in performance—is consistent with both hypotheses under consideration. The origin of the significant differences between the accuracy profiles for Tunes A and B is not clear at present.

There appears to be an inverse correspondence between the DAPs and the performance intensity profiles (Fig. 2b) in qualitative terms. However, the perception-performance correlations were small and nonsignificant: -0.23 for Tune A and -0.05 for Tune B. This was due to the absence of increased detection accuracy at the beginnings and ends of the tunes, where intensity was markedly reduced in performance. Also, the enormous increase in detection accuracy during the first subphrase has no correspondence in performance. Nevertheless, the two broad peaks in the accuracy profiles correspond to valleys in the performance intensity profile. It is difficult to conclude, therefore, that perception and performance are unrelated, but the parallelism certainly is less striking than in the case of timing.

The partial correspondence between the DAP and the performance intensity profile could have been mediated by bottom-up factors: either absolute pitch height, or the pitch difference between adjacent tones. Accuracy scores indeed correlated negatively with absolute pitch (Tune A: $r = -0.68$, $p < .001$; Tune B: $r = -0.49$, $p < .05$) and with directional pitch change (Tune A: $r = -0.52$, $p < .05$; Tune B: $r = -0.44$, $p < .05$). Intensity increments thus were more difficult to detect on higher-pitched tones, and/or when the pitch went up rather than down. Moreover, it can be seen in Figure 9 that accuracy scores *kept increasing* as the pitch went down (middle portions of first and second subphrases), whereas they were low and relatively stable when the pitch went up. In a stepwise regression analysis, however, pitch change made no significant contribution beyond the effect of absolute pitch height, which showed the stronger correlations with detection scores. The performance intensity profiles, however, had shown stronger correlations with pitch change than with pitch height. Thus pitch height seemed to be more of a factor in perception than in performance, whereas pitch change affected both. Effects of pitch height may be due to lower piano tones being perceived as louder than higher tones of equal intensity, due to their larger spectral bandwidth and slower decay times. This was confirmed in a recent study which required listeners to adjust the relative loudnesses of piano tones differing in pitch (Repp, 1995b), and it probably represents a genuine psychoacoustic effect that is not mediated by musical experience. Effects of pitch change, on the other hand, may reflect a more complex relational stimulus property of "pitch motion," which may or may not be purely bottom-up in nature (see the General Discussion below).

False alarm profiles. False alarm responses were much more frequent in this experiment than in Experiment 1.[18] The average percentage, relative to the number of target positions, was 24.8; relative to the number of misses (54.3%) it was 45.7%. That is, almost every other time that a target was missed, a false alarm response was given. Individual differences were considerable: Individual false alarm rates ranged from 5.2% to 47.7% (relative to the number of target positions). A few subjects actually had more false alarms than hits. (Note that this does not imply chance performance in the present paradigm.) Musicians gave somewhat fewer false alarms than the other subjects, but the difference was not significant. Surprisingly, however, there was a main effect of tune: False alarms were more frequent in Tune A than in Tune B [$F(1,17) = 4.71$, $p < .05$], just as in Experiment 1.

The aligned FAPs for the two tunes are shown in Figure 10. It can be seen that the profiles were quite similar, though response rates in some positions were elevated in Tune A relative to Tune B. The two sharp peaks on pitches $F^{\#}$ (second subphrase) and c (third subphrase) in Tune A correspond to downbeats. This cannot be evidence for a role of metrical structure, however, because the effect goes in the wrong direction: Downbeats should be expected to be louder, if anything, and hence should not attract false alarm responses. The most striking fact about the FAPs, of course, is their similarity to the DAPs shown in Figure 9. The correlations between these profiles were 0.69 for Tune A and 0.72 for Tune B, both $p < .001$. False alarms were frequent precisely in those positions where detection of intensity increments was easy. This is in accord with the top-down hypothesis, but it could also represent a psychoacoustically based bias. Since pitch height and pitch change have already been identified as relevant stimulus factors, it seems plausible that

the false alarm distribution reflects the operation of these factors as well. Their correlations with the FAPs were not impressive, however, due in part to the lower reliability of the FAPs, which reflected mainly the results of a few subjects with very high false alarm rates. Still, there were tendencies to give false alarms to low tones (Tune A: r = -0.30, n.s.; Tune B: r = -0.49, $p < .05$) and to tones lower than the preceding tone (Tune A: r = -0.13, n.s.; Tune B: r = -0.37, n.s.). The strong DAP-FAP correlation thus cannot be taken as unambiguous support for the top-down hypothesis; it is consistent with the bottom-up hypothesis as well.

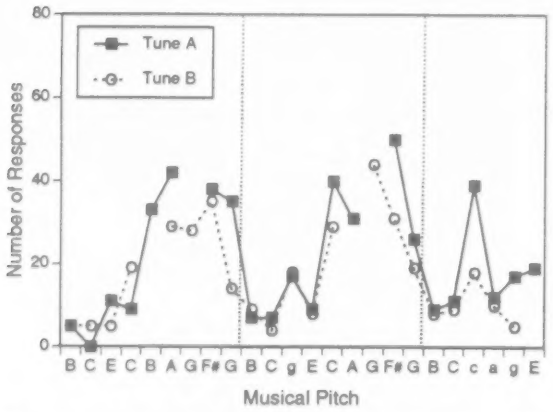


Figure 10. **Average false alarm profiles for Tunes A and B in Experiment 2, aligned according to pitch.**

Summary. As already suggested by Repp (1992b), it appears that detection of intensity increments is not governed by top-down expectations of performance microstructure to the same extent as may be the detection of duration increments. The perceptual results seem to reflect primarily a perceptual bias due to an interaction between pitch and perceived loudness, which may be specific to piano tones. The relation between perception and performance remains unclear in the case of dynamics.

GENERAL DISCUSSION

The present experiments explore a middle ground between traditional psychoacoustics and music perception. They deal with situations that are unusually complex and informal from a psychoacoustic perspective but unusually primitive and constrained from a musical perspective. Thus they are open to criticism from both sides. Yet it is necessary to begin to fill the yawning gap between these different research traditions in order to better understand the relevance of psychoacoustic findings to the perception of realistic music. The present research used melodies that, even though they were quite simple, lent themselves to expressive and aesthetically pleasing performance and could be listened to in a correspondingly musical mode. Thus they were not merely auditory sequences but also could be expected to call on listeners' aesthetic sensibilities and musical experience. The experimental task, although it was demanding and repetitive, also deliberately deviated from typical low-uncertainty psychoacoustic paradigms in order to get somewhat closer to realistic music perception.

To some extent, the very simplicity of the musical materials counteracted these efforts. However, the purpose of Experiment 1 was to attempt to replicate Repp's (1992a) findings on duration increment detection with materials that gave only limited room for purely stimulus-based (bottom-up) effects. The experiment was moderately successful in achieving this goal. Although neither the performance timing profile nor the DAP were as intricately structured as in the earlier study, due to the reduced complexity of the music, they were negatively correlated, and both were positively correlated with the FAP, suggesting an underlying bias. The auditory kappa effect could have provided an explanation for this bias, but there was no convincing evidence that this effect operated in the present melodies. In the absence of such a bottom-up explanation for the perceptual results and of a bottom-up rationale for a connection between perception and performance, the data lend support to the top-down hypothesis that listeners expected a certain timing microstructure, and that these expectations interacted with perception of duration increments. The absence of any effect of musical experience on the shape of the DAP or FAP is troublesome from that perspective, for it suggests that musically untrained people, too, have expectations about expressive timing, even though they have had little exposure to expressively played music. It is possible, however, that film music and the softer styles of popular music provide fairly universal exposure to the basic phenomena of expressive performance. The presence of corresponding effects of metrical structure in both perception and performance would have given a strong boost to the top-down hypothesis, but the general absence of clear metrical effects, attributed to the slow tempo and primitive rhythmical structure of the materials, is at least not inconsistent with the hypothesis.

The conclusions from Experiment 2 are different. It appears that perception of intensity increments is governed more by the bottom-up factors of pitch height and pitch distance than by any top-down expectations about dynamic performance characteristics. The absence of a strong perception-performance correlation, the obtained significant correlations of the perceptual results with the relevant stimulus factors, and the absence of effects of metrical structure and of musical experience are all consistent with a bottom-up explanation. This does not mean that top-down expectations about dynamics never play a role, only that they could not be convincingly demonstrated in this experiment, despite a highly varied performance intensity profile that provided a good basis for such expectations. In view of the strong DAP-FAP correlation, the observed bottom-up effects (greater detectability of intensity increments in low tones, and in tones following higher tones) probably represent variations in perceptual bias (directional) rather than variations in sensitivity (nondirectional). Subsequent studies bear more directly on this distinction (Repp, 1995c).

One reason for why perception of timing may be governed more by top-down processes than perception of intensity is that agogics provide a more important correlate of musical structure than do dynamics. Performers use a local slowing of tempo to mark structural boundaries, often in proportion to the depth of the boundary in the grouping hierarchy (Todd, 1985; Repp, 1992a). Dynamics, although they often are correlated with tempo modulations (Todd, 1992), may be only a secondary and less reliable indicator of structure. Since listeners' microstructural expectations presumably are driven by a structural representation of the music heard, they would then naturally carry much stronger biases with regard to timing than to dynamics. Performers, too, presumably have more freedom in varying the dynamic shape of a performance, without obscuring the structure in the process. Constraints on dynamics may arise primarily from the linear pitch sequence as such, rather than from a higher-level hierarchical representation.

For heuristic reasons, it has been attempted to draw a clear distinction between bottom-up (stimulus) and top-down (cognitive) factors in this research. However, this distinction can easily become blurred. For example, as already mentioned, whenever there is a consistent stimulus correlate of some perceptual effect (such as the apparent effect of pitch height on perceived loudness), the correlation can become part of listeners' expectations. Any bottom-up effect thus can simultaneously be a top-down effect, and other arguments (such as parsimony or even more elementary stimulus factors, such as the distribution of energy across critical bands and its effect on the auditory computation of loudness level) need to be invoked to establish the heuristic priority of the bottom-up account. Conversely, what seems like a top-down effect may actually reside in the stimulus structure, because this structure may be richer than has been granted. Ultimately, the division between bottom-up and top-down, like the analogous one between perception and cognition, is a *variable* determined by how much structure a theorist is willing to impute to the stimulus. The more structure is said to be in the world, the less needs to be attributed to the listener's cognitive processes and past experience; the structure is simply "picked up" rather than constructed or inferred (Gibson, 1966).

Jones (1976, 1987, 1990; Jones & Boltz, 1989) has been a prominent advocate of such a stimulus-structure-oriented approach to music. According to her, the pitch, loudness, and time dimensions of auditory patterns (including music) are "inextricably bound together and cannot be evaluated separately" (Jones, 1976, p. 329). More recently, she has proposed a two-component model of music performance which combines a rigid "vertical" timing component with a flexible "horizontal" component that generates expressive timing variation (Jones, 1987, 1990). The horizontal component "captures the shape and pacing of a melodic line, the velocity profiles of melodic phrases" and is controlled by a "motion generator" (Jones, 1990, p.227). Jones does not discuss *how* velocity profiles are aligned with the pitch structure of the music, but she leaves little doubt that the alignment is not arbitrary and to a large extent governed by the pitch structure of the melody within structural units ("melodic phrases"). Indeed, the notion that sequences of musical tones have inherent dynamic properties such as tension-relaxation or "tonal motion" has long been a staple of musicologists and philosophers of music (see, e.g., Truslit, 1938; Zuckerkandl, 1956; Repp, 1993b; Shove & Repp, 1995).

Such a view of music—as being invested with holistic and relational properties that go beyond the mere sequence of pitches and that appeal directly to the kinematics of the human body—provides a new perspective on the relationship between perception and performance that is at the heart of the present research. The traditional top-down and bottom-up hypotheses contrasted here

basically assume that a causal relationship underlies any observed perception-performance correlation: According to the top-down hypothesis, expectations about performance govern perception; according to the bottom-up hypothesis, perceptual distortions presumably underlie expressive performance strategies. Without the assumption of such causal connections, parallels between perception and performance cannot be explained by these hypotheses. However, an enriched view of the music itself makes it possible to see perception and performance as two parallel kinds of reactions (the listener's and the musician's) to the same complex information; hence the causality goes from the music to both, not from one to the other. Thus, for example, the gradual decline in the DAP towards the end of the musical excerpt as well as the performer's *ritardando* may both be reflections of the dissipation of tension in the music as it approaches its end, which is signalled both temporally and melodically (by the falling pitch and the harmonic cadence implied by the monophonic melody). In other words, the performer slows down because the music *asks for it*, and the listener "expects" the *ritardando* for the same reason. Of course, the musical structure does not completely determine perception and performance, but it exerts significant constraints on them, perhaps more so on perception than on performance, and—as the present research suggest—more on timing than on dynamics.

REFERENCES

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London: Academic Press.

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance* (pp. 76-181). Stockholm: Royal Swedish Academy of Music (Publication No. 39).

Collyer, C. E. (1974). The detection of a temporal gap between two disparate stimuli. *Perception & Psychophysics, 16*, 96-100.

Crowder, R. G., & NEATH, I. (1995). The influence of pitch on time perception in short melodies. *Music Perception, 13*, 000-000.

Dai, H., & Green, D. M. (1992). Auditory intensity perception: Successive versus simultaneous, across-channel discriminations. *Journal of the Acoustical Society of America, 91*, 2845-2854.

Divenyi, P. L. (1971). The rhythmic perception of micro-melodies: Detectability by human observers of a time increment between sinusoidal pulses of two different, successive frequencies. In E. Gordon (Ed.), *Experimental research in the psychology of music, Vol. 7* (pp. 41-130). Iowa City, IA: University of Iowa Press.

Divenyi, P. L., & Danner, W. F. (1977). Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception & Psychophysics, 21*, 125-142.

Divenyi, P. L., & Sachs, R. M. (1978). Discrimination of time intervals bounded by tone bursts. *Perception & Psychophysics, 24*, 429-436.

Drake, C. (1993). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society, 31*, 107-110.

Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics, 54*, 277-286.

Fitzgibbons, P. J., Pollatsek, A., & Thomas, I. B. (1974). Detection of temporal gaps within and between perceptual tonal groups. *Perception & Psychophysics, 16*, 522-528.

Formby, C., & Forrest, T. G. (1991). Detection of silent temporal gaps in sinusoidal markers. *Journal of the Acoustical Society of America, 89*, 830-837.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149-180). Orlando, FL: Academic Press.

Friberg, A. (1991). Generative rules for music performance: A formal description of a rule system. *Computer Music Journal, 15*, 56-71.

Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A major (K.331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-103). Stockholm: Royal Swedish Academy of Music.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Hirsh, I. J., Monahan, C. B., Grant, K. W., & Singh, P. G. (1990). Studies in auditory timing: 1. Simple patterns. *Perception & Psychophysics, 47*, 215-226.

Ivry, R. B., & Hazeltine, R. E. (1995). Perception and production of temporal intervals across a range of durations: Evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 3-18.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83*, 323-355.

Jones, M. R. (1987). Perspectives on musical time. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 153-175). Stockholm: Royal Swedish Academy of Music.

Jones, M. R. (1990). Musical events and models of musical time. In R. Block (Ed.), *Models of cognitive time* (pp. 207-240). Hillsdale, NJ: Erlbaum.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*, 459-491.

Jusczyk, P. W., & Krumhansel, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 627-640.

Krumhansl, C. L., & Jusczyk, P. E. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1*, 70-73.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Monahan, C. B., & Hirsh, I. J. (1990). Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics, 47*, 227-242.

Neff, D. L., Jesteadt, W., & Brown, E. L. (1982). The relation between gap discrimination and auditory stream segregation. *Perception & Psychophysics, 31*, 493-501.

Noorden, L. P. A. S. van (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology, The Netherlands.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 301-315.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception, 11*, 409-464.

Perrott, D. R., & Williams, K. N. (1971). Auditory temporal resolution: Gap detection as a function of interpulse frequency disparity. *Psychonomic Science, 25,* 73-74.

Repp, B. H. (1992a). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition, 44,* 241-281.

Repp, B. H. (1992b). Detectability of rhythmic perturbations in musical contexts: Bottom-up versus top-down factors. In C. Auxiette, C. Drake, & C. Gérard (Eds.), *Fourth Rhythm Workshop: Rhythm perception and production* (pp. 111-116). Bourges, France: Imprimérie Municipale.

Repp, B. H. (1993a). Some empirical observations on sound level properties of recorded piano tones. *Journal of the Acoustical Society of America, 93,* 1136-1144.

Repp, B. H. (1993b). Music as motion: A synopsis of Alexander Truslit's "Gestaltung und Bewegung in der Musik" (1938). *Psychology of Music, 21,* 48-72.

Repp, B. H. (1995a). Acoustics, perception, and production of *legato* articulation on the piano. *Journal of the Acoustical Society of America, 97,* 3862-3874.

Repp, B. H. (1995b). Relative loudness of piano tones differing in pitch. Manuscript in preparation.

Repp, B. H. (1995c). Variations on a theme by Chopin: Relations between perception and production of deviations from isochrony in music. Submitted for publication.

Shigeno, S. (1986). The auditory tau and kappa effects for speech and nonspeech stimuli. *Perception & Psychophysics, 40,* 9-19.

Shigeno, S. (1993). The interdependence of pitch and temporal judgments by absolute pitch possessors. *Perception & Psychophysics, 54,* 682-692.

Shove, P., & Repp, B. H. (1995). Musical motion and performance: Theoretical and empirical perspectives. In J. Rink (Ed.), *The practice of performance* (pp. 55-83). Cambridge, UK: Cambridge University Press.

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, 35A,* 377-396.

Sloboda, J. A. (1985). Expressive skill in two pianists: Metrical communication in real and simulated performances. *Canadian Journal of Psychology, 39,* 273-293.

Sundberg, J. (1988). Computer synthesis of music performance. In J. A. Sloboda (Ed.), *Generative processes in music* (pp. 52-69). Oxford, UK: Clarendon Press.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3,* 33-58.

Todd, N. P. McA. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91,* 3540-3550.

Todd, N. P. McA. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America, 97,* 1940-1949.

Truslit, A. (1938). *Gestaltung und Bewegung in der Musik.* Berlin-Lichterfelde: Chr. Friedrich Vieweg.

Williams, K. N., & Perrott, D. R. (1972). Temporal resolution of tonal pulses. *Journal of the Acoustical Society of America, 51,* 644-647.

Zuckerkandl, V. (1956). *Sound and symbol: Music and the external world.* Princeton, NJ: Princeton University Press.

FOOTNOTES

**Perception & Psychophysics,* in press.

[1]The sign of the correlation is due to the preferred representations of the data: percent correct rather than percent errors for the DAP, and IOI duration rather than local tempo for the performance profile.

[2]While the magnitude of the DAP-FAP correlation provides an indication of whether presumable bottom-up effects are due mainly to variations in bias or sensitivity, a better way of distinguishing these two possibilities is to compare increment and decrement detection. This was done in subsequent experiments (Repp, 1995c). The present paradigm does not lend itself to a signal-detection-theory analysis because the false-alarm rates are too unreliable and difficult to convert into the proportions needed for calculation of d' and beta.

[3]Similarly, the well-known demonstrations by Noorden (1975) and others of stream segregation as a function of pitch distance (see Bregman, 1990) typically involved faster rates of presentation than used here, simple rather than complex tones, and silent gaps between tones. Stream segregation is not likely to occur in the present paradigm.

[4]Results of a recent study do suggest that two piano tones are easier to compare with respect to loudness when they have the same pitch than when they differ in pitch (Repp, 1995b).

[5]Positions are denoted by a bar number followed by a beat number. It is assumed here that the subphrases have the same upbeat structure and the same length.

[6]Naturally, there were individual differences among the pianists in their timing and velocity profiles. These differences, which were more quantitative than qualitative, will not be discussed here in detail. The average profile is an estimate of what all performances have in common, and hence an estimate of what the average musical listener may expect to hear.

[7]Some individual pianists showed differences between the two profiles, but they were different in nature for each pianist and thus averaged out.

[8]The second criterion was relaxed for one subject who had had 12 years of piano and 9 years of violin instruction but did not currently play either instrument.

[9]Eight additional subjects were excluded because of difficulties in performing the task. Three were excused from the experiment because they could not hear any lengthening in the initial examples; of the five who completed the test, three responded randomly, one was close to chance and often gave more than two responses per trial, and one performed near chance during the first but not the second half of the test (which was highly atypical). Of these five subjects, three would have been classified as amateurs and two as nonmusicians. It is noteworthy that they ranged in age from 29 to 57; thus age may be a handicap in this task. Repp (1992b) similarly observed that musically untrained listeners often have great difficulty hearing timing differences.

[10]When an IOI is lengthened, the tone occupying it has extra time to decay before the next tone comes on, and the final decay following its nominal offset (the simulated key release) is correspondingly shortened (see Repp, 1995a). Thus there is less energy of the preceding tone at the onset and during the initial portion of the following tone, which may well enhance its perceived loudness.

[11]The superiority of the musicians would have been even clearer if one subject in that group had not performed rather poorly.

[12]In a forced-choice task, chance level would be about 1/7, considering that responses in three adjacent positions were accepted as correct. However, subjects responded only 16% of the time when they could not detect a target (see false alarms below), so the actual chance level was about $(1/7)^2$, or 2% correct.

[13]Starting with the last two tones in the first subphrase, every pair of successive IOIs in Tune A shows higher performance in the first (metrically strong) than in the second (metrically weak) position. (This is seen more clearly in Fig. 4.) However, since Tune B generally shows a similar (pitch-aligned) pattern (Fig. 5), it is difficult to attribute it to metrical alternation, which was just the opposite in Tune B than in Tune A.

[14]Incorrect responses could not simply be divided into false alarms and misses, because of occurrences such as a single-target trial with two false alarm responses, or with one correct response and one false alarm response.

[15]No subject had to be excluded because of inability to hear the intensity increments or because of random performance on the test, even though the task was quite difficult. (The same was true in Repp's, 1992b, Experiment 2.) It appears that intensity increment detection is a more straightforward task for untrained subjects than is duration increment detection.

[16]There was also a change in spectrum along with the intensity change, as the instrument used modelled the natural covariation between dynamics and spectral structure in piano tones. However, the resulting change in timbre (increase in brightness) was almost certainly too small to be detected as such.

[17]Given a basic guessing probability of about 1/7 and an average false-alarm rate of close to 50% (see below), the chance level in this experiment was roughly 7%.

[18]It seems unlikely that this higher incidence of false alarms was due to *random* variations in the perceived loudnesses of the different piano tones as a function of pitch (even though some random variability in peak intensity was found by Repp, 1993a). In that case, the FAP distributions (Figure 10) should have been flatter, and absolute detection accuracy should have been considerably worse than it was. The cause of the frequent false alarms probably was a *systematic* effect of pitch variation on perceived loudness.

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 193-209

Acoustics, Perception, and Production of *Legato* Articulation on a Digital Piano*

Bruno H. Repp

This study investigated the perception and production of *legato* ("connected") articulation in repeatedly ascending and descending tone sequences on a digital piano (Roland RD-250s). Initial measurements of the synthetic tones revealed substantial decay times following key release. High tones decayed faster than low tones, as they did prior to key release, and long tones decayed sooner than short tones because of their more extensive pre-release decay. Musically trained subjects (including pianists) were asked to adjust the key overlap times (KOTs) of successive piano tones so that they sounded optimally, minimally, or maximally *legato*. The results supported two predictions based on the acoustic measurements: KOTs for successive tones judged to be optimally or maximally *legato* were greater for high than for low tones, and greater for long than for short tones, so that auditory overlap presumably remained more nearly constant. For minimal *legato* adjustments the effect of tone duration was reversed, however. Adjusted KOTs were also longer for relatively consonant tones (3 semitones separation) than for dissonant tones (1 semitone separation). Subsequently, KOTs were measured in skilled pianists' *legato* productions of tone sequences similar to those in the perceptual experiment. KOTs clearly increased with tone duration, an effect that was probably motoric in origin. There was no effect of tone height, suggesting that the pianists did not immediately adjust to differences in acoustic overlap. KOTs were slightly shorter for dissonant than for consonant tones. They also varied with position in the ascending-descending tone sequences, indicating that the pianists exerted strategic control over KOT as a continuous expressive dimension. There were large individual differences among pianists, both in the perceptual judgment and in the production of *legato*.

This research was supported by NIH grant MH-51230. I am grateful to Charles Nichols for his extensive assistance, to Janet Hander-Powers and Nigel Nettheim for comments on an earlier draft, and to the participating pianists for their patience.

INTRODUCTION

A. Modes of articulation

On most musical instruments, successive tones can be produced in two basic modes of articulation: unconnected and connected. In the unconnected mode, perceptible intervals of (what seems to be) silence separate successive tones. On the piano, this is achieved by releasing a key before the next key is depressed. It is appropriate whenever the score indicates a rest or *staccato* articulation, and also at the ends of phrases or slurs as an aspect of "phrasing." The connected mode is generally referred to as *legato* articulation. Here the preceding tone seems to end at the same time as the following tone begins. Correspondingly, the pianist releases a key at about the same time as the next key is depressed.

The piano is one of a number of instruments that permit the simultaneous sounding of several tones. This is achieved by depressing several keys at the same time.[1] The possibility of this third mode of articulation, the simultaneous or chordal mode, has implications for *legato* articulation: In order to achieve a very smooth connection between tones, a pianist may release a key after depressing the key for the following tone, so that there is a small amount of overlap. The duration of this overlap (time of key release minus time of key depression for the following tone) will be referred to as *key overlap time* (KOT) in the following; it is positive when there is overlap, and negative when

the key release precedes the following key depression.[2]

A small amount of key overlap is not perceived as a simultaneity but rather as an increased connectedness of the tones. One reason why some amount of key overlap can be tolerated is that the sound level of a piano tone, after a rapid rise, decays as a function of time, so that the end of the preceding tone is usually much softer than the beginning of the following tone. Because of this discrepancy in relative intensity and the abrupt rise time of the following tone, masking may occur, so that a brief acoustic overlap may not be readily detectable.

However, masking is not likely to provide a full explanation because the acoustic overlap of successive piano tones is actually much more extensive than suggested by the KOT. Although the damper extinguishes the string vibrations when a key is released, this process is not instantaneous, and soundboard vibrations and acoustic reverberation in a room may further contribute to prolonging a tone's acoustic duration. The highest piano strings (usually starting with F#6) do not have any dampers at all. Thus, when two tones are perceived as unconnected (i.e., when KOT is negative), the apparent silent interval between them is at least partially filled by the decaying energy of the first tone, and there may in fact be some acoustic overlap. In typical *legato* articulation, where one key is released shortly after the depression of the next key, the acoustic overlap may be quite substantial, yet listeners do not complain about hearing simultaneous tones. This is probably due to auditory grouping of a tone's "tail" with its "body," and consequent perceptual segregation of the simultaneous tones, within limits (see Bregman, 1990).[3] Here we note only that a short KOT corresponds to a much longer *tone overlap time,* whose precise duration depends on the intensities and decay rates of the tones and on the ambient noise level.

These considerations give rise to a number of interesting questions about the acoustics, perception, and production of *legato* articulation on the piano--questions that the present study was intended to address in a preliminary way:

(1) *Acoustics*: What are the decay characteristics of piano tones, particularly after key release?

(2) *Perception*: How much key overlap (and consequent tone overlap) can listeners tolerate when judging the articulation of successive tones to be *legato,* before they start hearing simultaneities? Does experience as a pianist affect these judgments? What factors influence the amount of key overlap listeners are willing to tolerate?

(3) *Production*: What are the typical KOTs when pianists play *legato*? Is there variation among individual pianists in this respect? Do pianists vary their timing of key depressions to compensate for factors that affect perceptual overlap? Are there differences in degree of *legato* between pianists' right and left hands, and between pairs of fingers on each hand?

There is not much previous research addressing these questions; what little is known to the author is summarized in the following sections.

B. Acoustics

The acoustic decay characteristics of sustained (undamped) piano tones are fairly well understood (see, e.g., Martin, 1947; Weinreich, 1990; Wogram, 1990). After a brief rise time and early amplitude peak, the sound decays slowly, typically with an initial faster rate of decay giving way to a slower rate. These two decay rates seem to represent the respective contributions of vertical and horizontal string motion: The vertical component is initially larger but is transmitted to the vertically moving soundboard and hence decays to below the level of the horizontal component within a few seconds (Weinreich, 1990). The decay rate of the vertical component can be altered radically, however, by interactions among the two or three strings struck by the same key (Weinreich, 1990) and by the relative impedance of the soundboard at the vibrating frequencies (Wogram, 1990); thus the two components may not always be clearly distinguishable in the amplitude envelope. Tones above roughly 700 Hz (about F5) seem to have only a single decay rate (Martin, 1947). The fact most relevant to the present study is that the decay rate increases with fundamental frequency: While a low tone such as C2 takes about 4 s to decay by 20 dB (1/10 of the amplitude) on an upright piano, C4 takes only about 2 s, and C6 about 1 s (Wogram, 1990). Because of the complex resonance characteristics of the soundboard and other factors, however, the decay rates of tones close in pitch may differ substantially (Benade, 1990; Wogram, 1990). The precise decay characteristics of individual piano tones thus are largely instrument-specific and need to be measured in any particular experimental context.

The literature offers surprisingly little information about the decay characteristics of piano tones following key release, after the damper falls upon the strings. These *post-release decay times* are

likely to depend on the amplitude of string vibration when the damper touches the strings, on the thickness of the strings, on the weight and surface condition of the damper, and on the velocity of key release (damper lowering), among other things. Moreover, the post-release decay times of tones heard by a listener are probably a good deal longer than those of the strings alone, since they include the decay of (undamped) sound board vibrations and reverberation. Therefore, it is necessary to measure these times in any specific setting used for research purposes. In the first part of the present study, this was done on the output of an electronic instrument whose sounds presumably had been modeled on a specific piano recorded under specific conditions. The main concern here was not the representativeness of these measurements for pianos in general but an adequate characterization of the specific acoustic environment in which the following perception and production experiments were conducted.

C. Perception

In the large literature on auditory psychophysics, there does not seem to be a single study that investigated listeners' ability to detect the acoustic overlap between the end of one tone and the beginning of another tone. However, there are many studies on the detectability of a silent gap between two pure tones or noise bursts (e.g., Perrott & Williams, 1971; Williams & Perrott, 1972; Collyer, 1974; Fitzgibbons, Pollatsek, & Thomas, 1974; Divenyi & Danner, 1977; Neff, Jestaedt, & Brown, 1982; Shailer & Moore, 1983; Buus & Florentine, 1985; Formby & Forrest, 1991). Unfortunately, none of these studies employed complex tones with decaying amplitudes; amplitude envelopes were rectangular, and gap detection thresholds were generally on the order of a few milliseconds. Many studies have shown that the gap detection threshold increases as the frequency separation of two pure tones increases, and Dannenbring and Bregman (1978) have reported perception of illusory overlap when alternating nonoverlapping tones are sufficiently different in frequency to be allocated to separate auditory streams. The intensity of the sounds seems to have no effect on gap detection, except at very low levels. However, Plomp (1964) demonstrated in a widely cited study that the gap detection threshold increases as the relative sound level of the second of two noise bursts decreases. He attributed this finding to a decay of auditory sensation following the offset of the first noise. This internal decay, in terms of equivalent sound pressure level (dB), was a negative exponential function of time and reached the sensation threshold 225 ms after noise offset, regardless of noise intensity. That is, the internal decay time was constant, but the decay rate varied with sound level.

Plomp's results are cited here to acknowledge that a sound often does not end with its acoustic termination; if its acoustic offset is abrupt, it continues to resonate for some time in the listener's auditory system. When a sound decays over time, however, the physical input generally will determine the auditory sensation, since acoustic decay (in dB) generally is a linear rather than negative exponential function of time (Benade, 1990). Thus, unless the acoustic decay rate is very high, the sensation level due to internal decay of auditory input will generally be below the momentary input level. For this reason (and others), the psychoacoustic literature is not very helpful in predicting what degree of acoustic overlap of piano tones might lead listeners to judge them as connected (*legato*) or unconnected. Even a controlled psychoacoustic study with ramped complex tones and highly trained listeners would only provide a partial answer. The question is best addressed within the realistic sound environment of piano tones, using listeners with musical experience rather than extensive laboratory training.

A recent study by Kuwano, Namba, Yamasaki, and Nishiyama (1994) moved in that direction. In the initial, more psychoacoustic, part of their research, they synthesized a 13-note melody with sounds whose pressure envelopes were either rectangular or triangular (i.e., linearly decaying).[4] The melody was presented at a fixed tempo, and the overlap of the tones was varied by changing their total duration. Musically untrained subjects judged when the tones changed from "separated" to "connected," and from "connected" to "overlapping." The average tone overlap times corresponding to these two boundaries were -56 and 16 ms for steady-state tones but 30 and 104 ms for decaying tones. That is, listeners judged steady-state tones to be optimally connected when they had a brief silent gap between them, whereas decaying tones perceived as connected overlapped by 67 ms, on the average. Apparently, this overlap was not heard as such.

In the second, more realistic, part of their study, Kuwano et al. asked a pianist to play the tune on an electronic piano in eight different ways, ranging from "markedly separated" to "extensively overlapping," while keeping the tempo constant. The recordings were subsequently presented to listeners who judged them as "separated,"

"marginally connected," or "overlapping." Their responses confirmed the pianist's intentions: The performance intended to be "marginally connected" (i.e., *legato*) also received the highest number of judgments in that category. The average overlap time of successive tones in that performance was 240 ms.

Kuwano et al. calculated overlap times by reproducing the tones of the pianists' performances individually on a MIDI system and measuring the point at which each tone had decayed to 60 dB below its peak level; this point was taken to be the end of the tone, and its distance from the onset time of the originally following tone was the overlap time.[5] Although this -60 dB criterion is commonly adopted in measuring reverberation times and, according to Benade (1990), often corresponds to the threshold of audibility for decaying isolated sounds, it may be a rather liberal criterion for tones in context, as the final part of the decaying tone is probably masked by the following tone. Still, in the absence of an independent investigation of masking among piano tones, the estimate of acoustic overlap of *legato* tones provided by Kuwano et al. is the best currently available. Although the overlap times of tones judged "connected" were much shorter in their first experiment, for reasons that are not quite clear, their second experiment is more representative of actual piano performance.

D. Production

Key overlap times have been measured in several studies of piano performance, though none of them focused specifically on *legato* playing. Thus, Sloboda (1983) showed that pianists vary KOTs (i.e., articulation) to convey different metrical interpretations of the same sequence of notes.[6] In a follow-up study, Sloboda (1985) displayed frequency distributions of KOTs for two pianists playing two tunes. The distributions were bimodal, with one peak around 40 ms and the other somewhere between -100 and -220 ms, corresponding to *legato* and *staccato* modes of articulation, respectively.

KOTs were also measured by MacKenzie and Van Eerd (1990) in a study of rapid scale playing at several tempi. When the tempo was 8 tones per second or faster, average KOTs were 5 ms or less; the longest times (28 ms) were observed at the slowest tempo (4 tones per second). KOTs were about twice as long for the right hand as for the left hand; however, this difference was confounded with register, as the two hands played two octaves apart. Rapid scale playing is not conducive to optimal *legato* articulation, but neither does it permit *staccato*; thus these data may be representative of "minimal" *legato*.

Palmer (1989) reported KOTs, even though she defined overlap acoustically as "the overlapping time of two adjacent notes' amplitude envelopes" (p. 335). Since she used a synthesizer with tones that had a fixed 80 ms decay following key release, and since she averaged KOTs across a number of tones varying in articulation, her data are not very revealing from the present perspective. More relevant are average KOT profiles for 10 pianists playing simple tunes, presented by Drake and Palmer (1993). Maximum KOTs were between 0 and 50 ms, presumably reflecting *legato* articulation. This study also included an analysis of a concert pianist's performance of an excerpt from a Beethoven sonata, but the KOTs were not reported in detail. Interestingly, both Palmer (1989) and Drake and Palmer (1993) found that the KOT was shorter when one of the two tones, especially the first one, was relatively long.

The situation most conducive to *legato* playing, short of providing explicit instructions, is the expressive performance of a slow, melodious piece. Repp (1994, in press) selectively measured KOTs in two pianists' performances of Schumann's "Träumerei." One pianist (an amateur) generally showed KOTs around 0 ms, whereas the other pianist (a professional musician) showed surprisingly long KOTs, often of several hundreds of milliseconds. Since her performances did not sound abnormal, such extreme overlaps are evidently acceptable to the ear in an expressive performance of complex music, but they may represent an extreme case.[7]

E. The present study

The production data available so far suggest that *legato* articulation is most often associated with KOTs of 0-50 ms, though longer overlaps may be observed under some circumstances. None of these measurements were obtained in a situation in which pianists were instructed to play *legato*. Little attention has been given to acoustic tone overlap and to listeners' perceptual tolerance for such overlap, with the exception of the pioneering study by Kuwano et al. (1994). The present research took an approach similar to theirs, but it focused in more detail on the acoustic characteristics of the tones used and on several factors that may influence KOTs. The study included an acoustic analysis, a perceptual experiment, and a production experiment.

The purpose of the acoustic analysis was to describe the decay characteristics of the digital

piano tones used in the subsequent experiments. The analysis was expected to confirm the known fact that high piano tones decay more rapidly than low tones before the key is released. Therefore, the post-release decay times were expected to be shorter for high tones than for low tones. One question of interest was whether the post-release decay times reflect merely a tone's sound level at the moment of key release, or whether there are also differences between high and low tones in their post-release rates of decay.

In the perceptual experiment, the method of adjustment was used to obtain judgments of optimal and marginal *legato* from musically trained listeners, including pianists. The stimuli were scales and *arpeggi* varying along three dimensions: tempo, register, and interval step size (correlated in this instance with relative dissonance versus consonance of successive tones). It was predicted that more key overlap would be tolerated at slow tempi and in a high register, because long and high tones are softer at key release than are short and low tones, so that acoustic overlap is both less extensive and less audible. Listeners' sensitivity to these factors may reflect perceptual constancy in terms of some criterion of auditory tone overlap. It was also expected that more key (as well as acoustic) overlap would be tolerated with relatively consonant sequences of tones than with relatively dissonant ones, both for auditory and aesthetic reasons. Because of the higher commonality of partials of consonant complex tones (Kameoka & Kuriyagawa, 1969), their acoustic overlap may be less detectable as well as more pleasing to the ear than the overlap of dissonant tones.

In the production experiment, skilled pianists provided samples of their best *legato* by playing the scales and *arpeggi* used in the perception experiment. The question was whether they would adjust their KOTs in response to factors that affect acoustic overlap, or whether *legato* playing is motorically invariant and insensitive to these factors, even though they may influence perceptual judgments. (This hypothesis will be further qualified below.) Secondary questions concerned the existence of individual differences in KOTs, as well as possible differences between the left and right hands and among pairs of adjacent fingers.

I. DECAY CHARACTERISTICS OF THE DIGITAL PIANO TONES USED

The purpose of the acoustic analysis was to characterize the sound pressure envelopes of the piano tones used in the perception and production experiments, with particular attention to the decay following key release. The analysis was essential because the tones were produced by an electronic instrument about whose sound inventory no detailed specifications were available from the manufacturer. A previous study (Repp, 1993) had focused on their peak sound levels and on aspects of their spectral structure but had not considered their decay characteristics.

It would have led too far to investigate in detail the extent to which these synthetic piano tones were representative of natural piano tones. However, it seemed highly likely that they were modelled on a set of acoustically recorded tones. Thus, their decay probably represented not only the decay of string vibrations but also the decay of soundboard vibrations and acoustic reverberation in some enclosed space. This was as it should be because it corresponds to what a listener hears and because it makes the synthetic tones sound realistic, especially over earphones.

A. Method

The instrument was a Roland RD-250s digital piano using a proprietary "adaptive synthesis" algorithm. "Piano 1" sound was used with a constant MIDI velocity of 40, because this velocity was also used in the subsequent perception experiment.[8] To reduce the number of measurements, only every third tone was analyzed in the range from C2 (65 Hz) to C7 (2093 Hz). Four series of tones between these two endpoints were played under the control of a Macintosh IIvx computer, using the Performer MIDI sequencing program. The four series differed in the *nominal tone duration*, i.e. in the interval between MIDI "note on" (key depression) and "note off" (key release) commands, which was 250, 500, 750, or 1000 ms. The nominal intertone interval (between each MIDI "note off" command and the next "note on" command) was 500 ms.[9]

The analog output of the digital piano was input to a Macintosh IIci computer using AudioMedia software with a sampling rate of 44.1 kHz. The digitized waveforms were then analyzed using Signalyze software. The RMS amplitude envelope was computed for each tone using a window size of 30 ms. Since the program does not display dB values, all measurements were performed on the RMS amplitude values and subsequently converted into dB (with an arbitrary reference). As the output of the digital piano was deterministic and temporal resolution was very fine, it was not necessary to perform each measurement more than once. Apparent irregularities were double-checked, however.

B. Results and discussion

Figure 1 shows *pre-release decay* as a function of musical pitch, as measured in the 1000-ms tones. The graph shows each tone's peak RMS level, as well as the sound levels 250, 500, 750, and 1000 ms from energy onset.[10] Peak level was reached after a variable rise time, which ranged from 24 to 43 ms for C2 to C5 (except for A4, which had an unusually slow rise time of 89 ms) but was much shorter (around 5 ms) from $E^{b}5$ on.[11] Peak levels were fairly stable between C2 and C6, although there was some variation from tone to tone (as already demonstrated by Repp, 1993). Above C6, peak level decreased as a function of pitch. The initial pre-release decay also increased dramatically at high pitches. Tones below C6 decayed by only a few dB over the first 250 ms, whereas from C6 on there was a very substantial initial decay. Beyond 250 ms, the decay rates varied less dramatically as a function of pitch, except for the tones in the lowest octave, which decayed at a much slower rate. Some individual tones (e.g., $E^{b}4$, A5) had amplitude envelopes with irregular characteristics, which may reflect beats caused by slightly mistuned strings in the original piano that served as a model. The sound level of C7 beyond 500 ms was too low to be measured. It is clear from Figure 1 that the pre-release decay increased with pitch, as expected, though there were irregularities in this relationship which presumably reflect the complex acoustics of real pianos.[12]

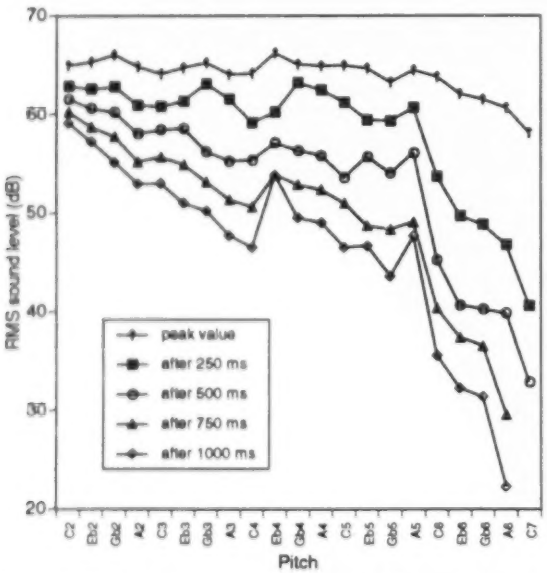


Figure 1. **Pre-release decay: Relative RMS sound levels of digital piano tones at their peaks and at four time points in their amplitude envelopes (measured from energy onset).**

The four lower functions in Figure 1 represent the sound levels at the time of key release for the four tone durations employed here. Since this sound level decreased as pitch increased and as tone duration increased, the *post-release decay time* must likewise have decreased with increasing pitch and increasing duration. This is confirmed in Figure 2, which shows how soon after key release tones of different nominal durations reached 1/10 (-20 dB) or 1/100 (-40 dB) of their peak amplitude.[13] Missing data points indicate that the specified level was reached before key release. The post-release decay times were quite substantial. Even by the conservative -40 dB criterion, the lowest tones took about 300 ms to decay, and tones up to C6 took at least 100 ms. Tones above C6 decayed somewhat sooner, though A6 and C7 were abnormal in that they showed much slower decay when released early than when released late. The reason for this anomaly was not clear, as these tones should not have been affected at all by key release because of the absence of dampers for the highest strings in a real piano.

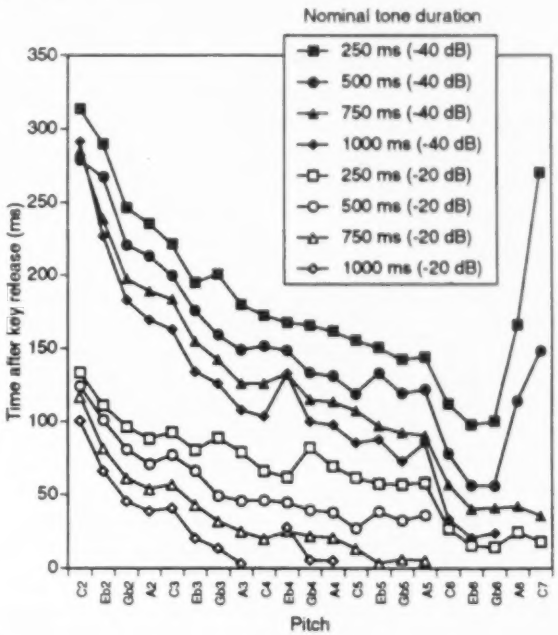


Figure 2. **Post-release decay times: Time after key release by which the tones had decayed to -20 and -40 dB of their peak level.**

The effects of pre-release decay on the post-release decay time, shown in Figure 2, would have been obtained even if the *post-release decay rate* had been constant. However, higher-pitched tones also decayed faster than lower tones, not only

before but also after key release. Pre-release tone duration, on the other hand, did not seem to have any systematic effect on post-release decay rate; therefore, Figure 3 shows the data averaged over the four nominal tone durations. The figure shows the time it took for the post-release sound level to decay by 20 dB. (This is the time difference between the -20 dB and -40 dB points in Figure 2.) This time decreased from about 180 to 90 ms over the pitch range investigated (i.e., the decay rate increased from about 11 dB/cs to 22 dB/cs), and the decrease as a function of pitch was much steeper during the lowest octave than during the higher octaves. The points for the two highest pitches have been omitted in the figure because of their much slower post-release decay rates (cf. Figure 2). The lines were fitted by hand to indicate the general trend of the data; again, there are some irregularities.[14]

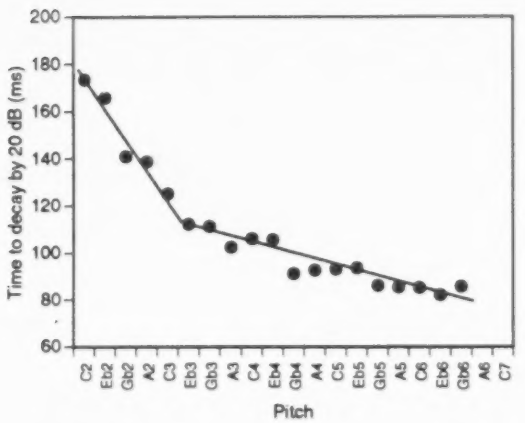


Figure 3. **Post-release decay rates: The time it took for tones to decay by 20 dB following key release.**

Although there seem to be no data in the literature to compare the present results with, the complex and somewhat irregular acoustic characteristics of the present tones suggest that they were indeed modelled after acoustically recorded piano tones. It should be noted, however, that natural piano tones exhibit much greater variation in peak sound level (Repp, 1993); some kind of equalization must have been applied in the proprietary synthesis scheme that generated the digital tones. To what extent the present findings are representative of the decay characteristics of natural piano tones remains to be determined. However, they adequately describe the acoustic environment within which the following experiments were conducted.

II. PERCEPTUAL JUDGMENT OF *LEGATO* ARTICULATION

The purpose of this experiment, as already indicated, was to investigate the influence of three factors (register, tempo, and relative consonance) on the amount of key overlap perceived as *legato*. An adjustment task was used to determine the overlaps judged to represent the "best" as well as "minimal" and "maximal" *legato*. It was expected that listeners would tolerate more key overlap in conditions where there is less acoustic overlap due to shorter post-release decay times, namely in the high register and at a slow tempo (long tone durations). Furthermore, it was predicted that more key (and acoustic) overlap would be tolerated for relatively consonant than for dissonant tones. (The consonant tones were also more widely separated in pitch, which could lead to the same prediction.) Half of the musically trained subjects were pianists, and it was of interest whether their specific experience would be reflected in different criteria for, and/or reduced variability of, their adjustments.

A. Method

1. Subjects. Fourteen paid volunteers participated. All but one were Yale undergraduates; they ranged in age from 18 to 25. All subjects were musically trained, having received between 8 and 15 years of formal instruction on at least one instrument. Seven were pianists (though several of them played a second instrument as well); the others played various instruments including violin, cello, double bass, oboe, trumpet, and guitar.

2. Materials and procedure. An interactive adjustment task was set up using the Roland RD-250s digital piano interfaced with a Macintosh IIvx computer. The control program was written using the Max graphic programming environment. The program created various random orders of 24 tone sequences resulting from the combination of three registers, four tempi, and two step sizes (or degrees of relative consonance). All tones had a constant MIDI velocity of 40. Each sequence consisted of a continuously ascending and descending scale or arpeggio based on five different tones. The three registers (low, medium, high) represented starting frequencies of C2 (65 Hz), C4 (262 Hz), and C6 (1047 Hz), respectively. The four tempi represented tone inter-onset

intervals (IOIs) of 260, 519, 779, and 1039 ms.[15] The two step sizes were 1 and 3 semitones (st), so that the tone sequences represented either a short chromatic scale extending over 4 st (a major third) or a diminished-seventh-chord arpeggio extending over one octave. Tones separated by 1 st formed the highly dissonant interval of a minor second, whereas tones separated by 3 st formed the moderately consonant interval of a minor third. Sequences were started and stopped by clicking START and STOP "buttons" on the computer screen. Nominal tone duration (key release time, and hence also KOT) was controlled by a horizontal "slider" on the screen that could be dragged or clicked with the mouse. Each sequence started with the slider in the left-most position, corresponding to a nominal tone duration of 150 ms, which made the tones sound definitely unconnected. Nominal tone durations controlled by the slider ranged from 150 to 1500 ms in 10-ms steps; they were not displayed numerically.

Subjects sat at the computer and listened binaurally to the output of the digital piano over Sennheiser HD540II earphones. The volume was set at the same comfortable level for all subjects. After receiving written instructions and a few minutes of free practice, subjects completed four blocks of 24 trials each, with short breaks in between. Each block was initiated by the experimenter who reset the program, which then generated a new random sequence of the same 24 trials. The current trial number was displayed on the computer screen. The subject initiated each trial by clicking the START button and terminated it by clicking the STOP button after adjusting the slider according to the criterion specified. The program stored the slider settings, and the experimenter saved them in a file at the end of each block. Each block took between 10 and 15 minutes to complete.

Instructions were different for each of the first three blocks. In the first block, the subject was asked to adjust the slider so as to find the "best" *legato*. (S)he was advised to move the slider slowly to the right until the tones sounded not only connected but unacceptably overlapping, and then to reverse direction and try to "zero in" on the optimal *legato* setting by moving the slider back and forth over a narrower region. (S)he was warned not to "overshoot" the target zone on the slider when the tempo was fast.[16] In the second block, the subject was asked to zero in on the boundary between unconnected and connected tones, and to find the setting that was just barely acceptable as *legato* ("minimal" *legato*). In the third block, the subject was asked to find the highest (right-most) setting that was still acceptable as *legato*, before the tones started to sound noticeably overlapping ("maximal" *legato*). The fourth block replicated the first, the task being again to find the best *legato*.

3. Data analysis. KOTs were determined by subtracting the IOI durations from the nominal tone durations adjusted by the subjects. In a few instances of overshoot, where the KOT exceeded the IOI (indicating that the subject was oblivious to the complete overlap of successive tones), the IOI was subtracted once again, yielding the KOT of tones separated by one intervening tone. Twelve missing data points (the program occasionally failed to save the last trial in a block) were replaced by the group mean for that particular condition. Repeated-measures analyses of variance were conducted on the data from each block, with the three independent variables (IOI, register, step size) as within-subject factors, and with pianistic expertise as a between-subject factor.

B. Results and discussion

The results are summarized in Figure 4. Consider first the "best" *legato* adjustments, which are shown averaged over blocks 1 and 4. There was substantial between-subject variability. (Note that standard errors are shown, not standard deviations.) The variability increased with register, step size, and IOI. Nevertheless, there were a number of reliable effects. First, as predicted, KOT increased with register [$F(2,24) = 25.52$, $p < .0001$]; the average KOTs were 14, 62, and 139 ms, respectively. Second, also as predicted, KOT increased as a function of step size or relative consonance [$F(1,12) = 20.96$, $p < .0007$]; the averages were 48 and 95 ms for 1- and 3-st step sizes, respectively. The effect of step size was reduced in the high register, which was reflected in a two-way interaction [$F(2,24) = 3.65$, $p < .05$]. Third, there was also a main effect of IOI [$F(3,36) = 3.57$, $p < .03$], though it was smaller and nonmonotonic; the respective average KOTs were 49, 84, 83, and 70 ms. The nature of the IOI effect varied with both register and step size; the triple interaction was significant [$F(6,72) = 2.43$, $p < .04$]. Two additional significant effects were the main effect of blocks [$F(1,12) = 6.59$, $p < .03$], due to shorter KOTs (by 20 ms) in block 4 than in block 1, and the block by step size interaction [$F(1,12) = 5.42$, $p < .04$], due to a reduced step size effect in block 4. Variability was also reduced in block 4. Pianistic expertise had no effect at all.

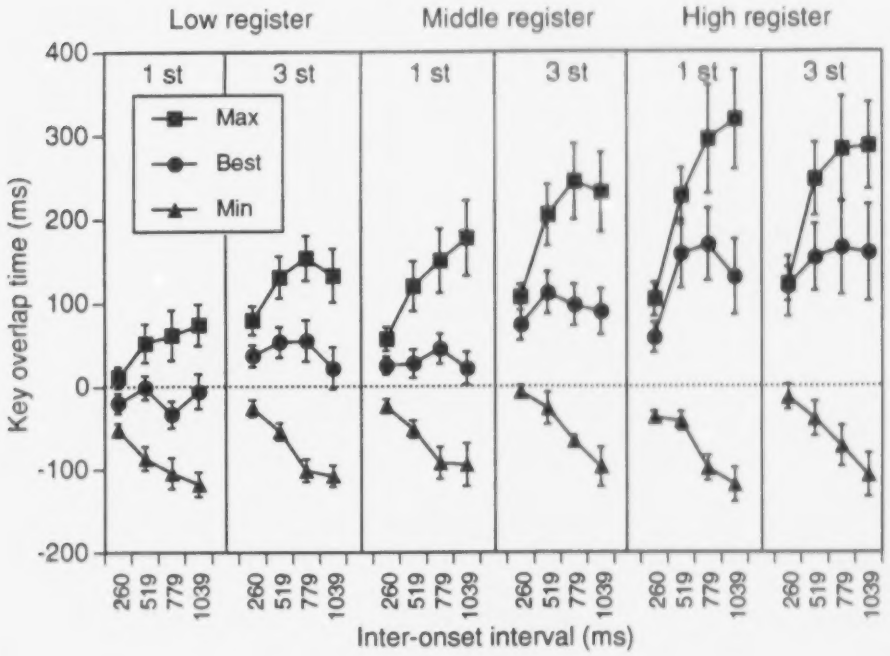


Figure 4. **Adjusted KOT in three conditions as a function of register, step size, and IOI. The bars represent plus/minus one standard error.**

The results for "maximal" *legato* judgments were basically similar, except that the adjusted KOTs were longer and the effect of IOI was stronger and more nearly monotonic. Variability was very high and increased with register and IOI, though not with step size. The effect of register was pronounced [$F(2,24) = 21.45$, $p < .0001$], with average KOTs of 87, 162, and 235 ms, respectively. The predicted effect of step size (or consonance) was present [$F(1,12) = 8.91$, $p < .02$], though not very large; the average KOTs were 137 and 185 ms, respectively. The step size effect was again absent in the high register, although the two-way interaction fell short of significance. The effect of IOI was highly significant [$F(3,36) = 19.20$, $p < .0001$], due to a negatively accelerated increase in KOT with IOI (80, 164, 198, and 204 ms, respectively). There was a significant IOI by register interaction [$F(6,72) = 3.71$, $p < .003$], due to larger effects of IOI as register increased. There was no effect of pianistic expertise.

Adjustments of "minimal" *legato* exhibited much less variability. The average KOTs were negative, indicating that nominal gaps of up to 100 ms can still be acceptable as *legato*, depending on the condition. The most striking effect here was that of IOI [$F(3,36) = 25.22$, $p < .0001$], though it was inverted: KOT *decreased* (i.e., nominal gap time increased) as IOI increased, the average durations being -26, -50, -89, and -107 ms, respectively. There was also a significant effect of step size [$F(1,12) = 9.40$, $p < .01$] in the predicted direction, with average KOT being less for chromatic scales (-76 ms) than for diminished-seventh-chord *arpeggi* (-60 ms). Finally, there was an effect of register [$F(2,24) = 3.85$, $p < .04$], though it was not monotonic: KOTs were shortest for low tones (-81 ms) and longest for medium-pitched tones (-57 ms), with high tones in between (-66 ms). No other effects reached significance.

From Figure 4 it is clear that the *range* of KOTs acceptable as *legato* increases dramatically with IOI and with register. Since several degrees of connectedness are probably discriminable within the larger ranges (though the relevant perceptual experiments remain to be conducted), the results imply that, the slower the tempo and the higher the register, the greater the variety of possible *legato* nuances. The opposite effects of IOI on the upper and lower boundaries of the *legato* range probably account for the irregular effect of IOI on

BEST COPY AVAILABLE

"best" *legato* judgments, which lie approximately in the center of the range in the low and middle registers, but closer to the upper boundary in the high register.

The effect of IOI on "maximal" *legato* judgments was as expected, with increasing KOTs being tolerated as IOI increased. This is almost certainly due to the lower sound levels at release and the shorter post-release decay times of long tones, which result in reduced acoustic and auditory overlap with the following tone. It is noteworthy that the IOI effect was largest between 250 and 500 ms and smallest or absent between 750 and 1000 ms. This agrees with the faster initial decay of piano tones, which takes place during the first 500 ms or so (see Figure 1).

The strong effect of register for both "best" and "maximal" *legato* judgments was also in the predicted direction, with the largest KOTs in the high register and the smallest in the low register. This is consistent with the much faster decay of high than low tones, both before and after key release. Unlike the effect of IOI, the register effect was of similar magnitude in "best" and "maximal" *legato* judgments.

The predicted effect of step size was obtained for both "best" and "maximal" *legato* judgments, but it was virtually absent in the high register. This interaction is compatible with an interpretation in terms of relative consonance. The relative dissonance of complex tones has been attributed to the interaction of individual partials (Plomp and Levelt, 1965; Kameoka & Kuriyagawa, 1969), and since high piano tones have fewer significant partials than low tones, they are likely to show fewer such interactions and hence smaller effects of perceived dissonance. It is difficult to see how such an interaction could have arisen from fundamental frequency separation alone. The fundamentals of tones separated by 3 st were well within the auditory filter bandwidth (Moore & Glasberg, 1983) in the low register but were separated by more than one critical band in the high register, whereas frequencies separated by 1 st were always within the same critical band. If anything, this should have led to a larger step size effect in the higher register.

The "minimal" *legato* judgments essentially represent gap detection thresholds. Of course, these estimates are much less precise than those obtained in typical psychoacoustic experiments, but their ecological validity may be greater. It is interesting to note that the average adjustments never corresponded to a key overlap, not even for high tones whose rapid pre-release decay caused a substantial drop in sound level before the onset of the following tone. Moreover, even though this drop increased with tone duration, more rather than less of a nominal gap was needed to hear a separation between long tones, and this was true regardless of pitch, the effect of register being rather small and nonmonotonic. Thus it was not the case that a sufficiently large drop in the amplitude envelope disrupted perception of connectedness. Rather, it seemed as if long tones sounded inherently more *legato* than short tones, so that more of a physical separation was required to hear them as disconnected. Although Kuwano et al. (1994) did not investigate the effect of nominal tone duration on perceptual judgments, their measurements of acoustic overlap times in a pianist's production show longer overlaps for long than for short tones when the intention was to play with various degrees of connectedness or overlap, but also longer acoustic gaps for long than for short tones when the intention was to play in a disconnected mode.[17] This pattern seems congruent with the present perceptual findings of an increased range of KOTs for long tones and of an inverted effect of tone duration (i.e., IOI) on nominal gap durations.

Kuwano et al. (1994) did not report KOTs but only acoustic overlap times, based on a -60 dB criterion for the end of a tone. According to that measure, the acoustic overlap was less than 170 ms when listeners gave predominantly "separated" responses, about 240 ms when they judged the tones to be "marginally connected," and more than 280 ms when they gave mostly "overlapping" responses. The IOIs were 600 and 300 ms (Kuwano, pers. comm.; the test melody contained both quarter and eighth notes), the pitch steps varied between 2 and 5 st, and the pitches ranged from F4 to F5. The present 260 and 519 ms IOI conditions in the middle register with a 3 st step size come closest to their stimuli. The average post-release decay time to -60 dB (extrapolated from the -20 and -40 dB points in Figure 2) of the relevant tones is roughly 250 ms. This implies acoustic overlaps of about 230 ms for "minimal" *legato*, 340 ms for "best" *legato*, and 400 ms for "maximal" *legato*. If the "marginally connected" category of Kuwano et al. is equated with the present "minimal" *legato*, then the data seem in agreement. It seems, however, that the present "best" *legato* stimuli would have been judged by their listeners as "overlapping," and the present "maximal" *legato* stimuli, as "extensively overlapping." There are a number of differences between the studies, however, that could account

for this apparently lower tolerance for overlap in their subjects, such as their use of musically untrained subjects. Also, their piano tones may have had decay characteristics different from those of the present tones; unfortunately, they did not describe those characteristics.

On a more general level, the present data agree with their findings in demonstrating that a substantial acoustic overlap can be tolerated by listeners before they complain about simultaneity of pitches. However, the results should not be taken to imply that the overlap is not detectable as such, even though the end of the "tail" of the decaying tone is almost certainly masked by the more powerful following tone. To assess the auditory detectability of overlap and the masking between simultaneous complex tones, precise psychoacoustical experiments are required. What matters more than detectability in a musical context, however, is perceptual and aesthetic tolerance within a specific instrumental environment and an associated performance tradition.

Because of masking between and/or perceptual segregation of simultaneous tones, it seems unlikely that any perceptual criterion (either "best" or "maximal" *legato*) corresponds to a fixed amount of acoustic overlap. This possibility was explored briefly by adding the post-release decay times determined in the first part of this study to the KOTs found in the second part. As expected, there was no constancy overall, though low-pitched tones judged to be maximally *legato* all overlapped by about the same amount. The listeners' responses in the adjustment task may be taken as *prima facie* evidence for some *auditory* constancy in terms of degree of connectedness. However, there may be no simple acoustic correlate of this constancy.

A final word is in order about individual differences. There was no difference between experienced pianists and nonpianists, but there was large variability within each group. The most unusual subject was a guitarist whose adjusted KOTs were all negative, even for "maximal" *legato*. Apparently, he was acutely sensitive to acoustic tone overlap and employed a criterion appropriate for his own instrument. Of course, he contributed strongly to the variability among the nonpianists. But even when his data were excluded, there was no clear difference between the two groups, and pianists themselves apparently had widely divergent criteria for what counted as a good *legato*. This may be related to individual differences in average KOT during *legato* articulation, an aspect of a pianist's "touch." The final part of this study investigated this production aspect of *legato* playing.

III. PRODUCTION OF *LEGATO* ARTICULATION BY PIANISTS

The purpose of this experiment was twofold. One aim was to measure the KOTs pianists produce when they intend to play optimally *legato* and to assess the magnitude of individual differences, as well as possible differences between right and left hands and between pairs of fingers. The second aim was to investigate whether pianists adjust, consciously or subconsciously, to factors that affect acoustic tone overlap and perceptual judgments of *legato* style, as demonstrated in the previous experiment. The pianists were asked to play scales and *arpeggi* like those used as stimuli in Part II, on the same instrument. Thus they were operating under similar acoustic conditions, and even though the sounds were synthetic and the keyboard felt different from that of a real piano, it was expected that, if adjustments in *legato* playing occur in response to acoustic factors on a real piano, they would also occur on a digital piano.

The same three factors as in Part II were varied in the materials, and the predictions were the same: Pianists were expected to show longer KOTs in a high register than in a low register, at a slow tempo (long IOIs) than at a fast tempo (short IOIs), and in a relatively consonant than in a dissonant sequence of tones. A complicating circumstance, however, was that these factors, tempo and step size in particular, may also have purely motoric consequences that are independent of the auditory feedback about tone overlap. Thus it seems that a smooth *legato* is more difficult to achieve (and perhaps also aesthetically less desirable) at a fast than at a slow tempo; if so, this effect reinforces the prediction based on acoustic considerations, but results supporting the prediction then cannot be attributed to a single cause. The situation is different with regard to step size: It may be more difficult to achieve a smooth *legato* when the fingers are spread (3 st step size) than when they are close together (1 st step size); if so, this effect counteracts the predicted effect of relative consonance, so that attribution of an obtained effect to acoustic-aesthetic or motoric causes is possible. Only a change of register seems to have no obvious motoric implications, as long as each hand stays within its typical range on the keyboard. Thus an

effect of register on KOT would provide the best evidence for an adjustment to acoustic conditions.

Individual differences among pianists in average KOT were of interest because they may reflect an aspect of the elusive quality of "touch." Since in much music the right hand plays the melody and the left hand the accompaniment, it was also considered possible that legato style is better developed in the right hand, leading to longer KOTs. As noted in the Introduction, MacKenzie and Van Eerd (1990) observed such a hand difference in rapid scale playing, but the fact that the right hand played in a higher register was not controlled for. In the present design, to avoid awkwardness, the low register was played only with the left hand, and the high register only with the right; however, the middle register was assigned to either hand, thus making a direct comparison possible. Finally, it was hypothesized that there might be more overlap between the "weak" fourth finger and its neighbors than between the more independent first three fingers.

A. Method

1. Subjects. The subjects were eight highly skilled pianists. Four of them (all female) were graduate students at the Yale School of Music; three (all male) were seniors in Yale college and, as winners of the annual undergraduate concerto competition, had performed as soloists with the Yale Symphony Orchestra; and one (female, older than the others) was a semi-professional accompanist. They were paid for their participation.

2. Materials. The materials were very similar to those used in Experiment 1, except that they were shown in musical notation on separate sheets of paper. Each scale or arpeggio ascended and descended three times and ended on a long note. The step size was 1 or 3 st, and the starting pitch was C2, C4, or C6. Two identical versions were prepared for the middle register, one marked "with the left hand" and the other "with the right hand." The low register examples were to be played with the left hand, and the high register examples with the right hand. Three tempo conditions were indicated by the note values: sixteenth notes, eighth notes, or quarter notes. All examples were to be played at about 60 beats per minute, so that the tempi corresponded to IOIs of approximately 250, 500, and 1000 ms, respectively.

3. Procedure. The subject sat at the Roland RD-250s keyboard and listened to its output over Sennheiser HD540II earphones. A metronome flashing silently at 60 beats per minute stood within view. After presenting written instructions and permitting a brief warm-up period on the instrument, the experimenter placed the randomly shuffled 24 music sheets before the pianist, one at a time. (S)he was asked to play each example "with (her or his) best legato" at the tempo indicated but not exactly with the metronome; expressive timing and dynamics, to the extent that the simple materials encouraged them, were welcome. The overall dynamic level asked for was mezzoforte. The fingering was prescribed in the instructions as 5-4-3-2-1-2-3-4-5 for the left hand and 1-2-3-4-5-4-3-2-1 for the right hand.[18] The experimenter monitored the playing of all examples to make sure that the instructions were followed. If either the subject or the experimenter was not satisfied with an example, it was repeated immediately. All productions were recorded in MIDI format using Performer software.

4. Data analysis. KOTs were calculated and subjected to repeated-measures analyses of variance. Three separate analyses were conducted: on the left-hand data, on the right-hand data, and on the middle-register data for both hands. The factors in these analyses were step size (2 levels), IOI (3 levels), register or hand (2 levels), fingers (8 levels), and repetitions (3 levels), with subjects as the random factor yielding the interactions that served as error terms. Note that the fingers factor (i.e., pairs of fingers: 5-4, 4-3, 3-2, 2-1, 1-2, 2-3, 3-4, 4-5 for the left hand, and 1-2, 2-3, 3-4, 4-5, 5-4, 4-3, 3-2, 2-1 for the right hand) was not decomposed into finger pairs (4 levels) and order of fingers within each pair (2 levels), for reasons that will become evident. The repetitions factor was treated as crossed with the other factors, not as nested within examples. Analogous ANOVAs were also performed on each individual pianist's data, in which case repetitions served as the random factor.

B. Results and discussion

The main results, averaged across pianists, repetitions, and fingers, are shown in Figure 5. The average KOTs ranged from about 50 to 150 ms across the different conditions. The most striking effect was that of IOI, which was significant in all three analyses [left hand: $F(2,14) = 7.93$, $p = .005$; right hand: $F(2,14) = 9.68$, $p < .003$; both hands: $F(2,14) = 9.47$, $p < .003$]. KOTs increased as the tempo decreased, as was predicted on both acoustic and motoric grounds.

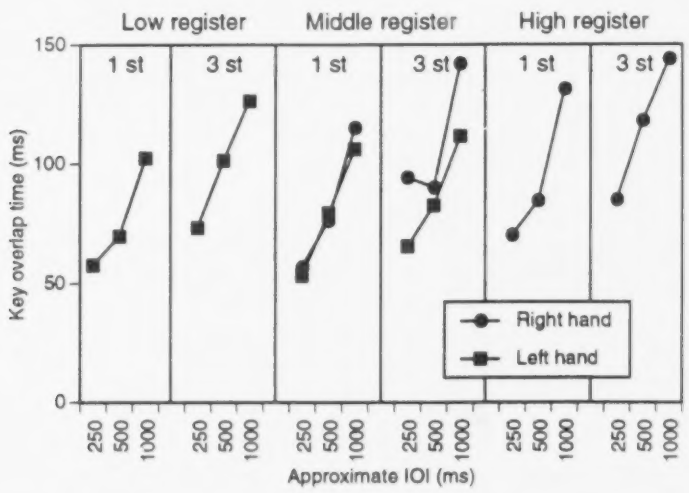


Figure 5. **Legato production: Average KOT as a function of IOI, step size, register, and hand.**

There was also an effect of step size or relative dissonance: KOTs were slightly longer for diminished-seventh-chord *arpeggi* than for chromatic scales [left hand: $F(1,7) = 4.14$, $p < .09$; right hand: $F(1,7) = 21.76$, $p < .003$; both hands: $F(1,7) = 15.11$, $p = .006$]. The effect was reliable only for the right hand, and in the middle register there was a significant interaction between hand and step size [$F(1,7) = 5.71$, $p < .05$]. Since there is no obvious motoric reason why *arpeggi* should be played more *legato* than chromatic scales, the step size effect probably reflects an adjustment to the relative consonance or dissonance of the successive tones. The greater sensitivity of the right hand to this dimension may be due to its leading role in melodic material.[19]

There was no significant effect of register for either hand. Although it seems that more overlap occurred in the high than in the low register, this difference may be due to hands rather than register. The absence of a register effect indicates that the pianists did not adapt their *legato* technique to the acoustic decay characteristics of the tones. The effect of IOI may then also be motoric rather than perceptual in origin.

There was a significant effect of hand in the middle register [$F(1,7) = 6.82$, $p < .04$]. KOTs were longer for the right hand (101 ms average overall) than for the left hand (86 ms). Although seven of the eight pianists showed a difference in this direction, it was individually reliable for only three. Note also that, in the middle register, the hand difference was apparently restricted to the *arpeggi*.

As expected, there were striking individual differences in average KOTs. Individual grand averages ranged from 27 ms to 145 ms. There was no obvious relation to either gender or relative experience of the pianists. Although experienced pianists can undoubtedly change their degree of *legato* when the music requires it, the fact that they produced such different KOTs in the same experimental situation suggests genuine individual differences in *legato* technique or "touch."[20]

The final effect to consider is that of fingers, which is portrayed in Figure 6. It was highly reliable [left hand: $F(7,49) = 7.86$, $p < .0001$; right hand: $F(7,49) = 13.15$, $p < .0001$; both hands: $F(7,49) = 12.93$, $p < .0001$] and was shown by every pianist. Its pattern was not what had been expected, however. Rather than reflecting longer KOTs for less independent pairs of fingers, it represented an interaction between finger pairs and order or, more simply, a main effect of position in the scale or *arpeggio:* In each upward movement *and* in each downward movement, KOTs started out long and then decreased until the scale or *arpeggio* reversed direction. This decrease was fairly linear in scales but apparently nonlinear in *arpeggi*. The finger by step size interaction was significant for the right hand [$F(7,49) = 4.10$, $p < .002$] and for both hands [$F(7,49) = 7.68$, $p < .0001$]; for the left hand, the triple interaction with register was significant instead [$F(7,49) = 3.36$, $p < .006$]. Although the two hands are shown separately in Figure 6, the interaction of fingers with hand was not significant.

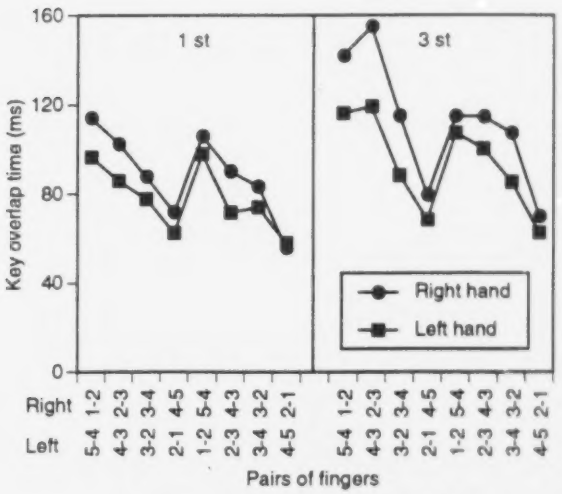


Figure 6. **Legato production: Average KOT as a function of finger pair, step size, and hand.**

The consistency of the finger effect indicates that it represents an important aspect of *legato* articulation, though its exact cause is uncertain at present. One possibility is that the forearm rotates as the scale or *arpeggio* is being played, transferring weight from one part of the hand to the other; this larger-scale movement may lag behind the fingers, causing decreased key overlap when moving in a given direction. Another factor that could contribute to the relatively short KOT just before a reversal is that the same finger must be used twice in close succession (2-1-2 or 4-5-4). However, there was no finger by IOI interaction, which would be expected if anticipatory lifting of a finger played an important role. Another possible factor that apparently can be ruled out is dynamic variation. The pianists naturally tended to increase the dynamic level as they went up and to decrease it as they went down a scale or *arpeggio*. While the absolute and especially the relative sound levels of successive tones could have an influence on the perception and production of KOTs, note that the obtained parallel trends for the ascending and descending halves (Figure 6) is contrary to the opposite trends expected on the basis of dynamics. Most likely, therefore, the finger effect, like the tempo effect, is not perceptual but rather motoric or cognitive in origin.

IV. GENERAL DISCUSSION

The present study combined acoustic analyses, perceptual judgments, and production measurements in an attempt to obtain some basic information about *legato* playing. Stimulated by the recent demonstration by Kuwano et al. (1994) that tones played and judged to sound connected show considerable acoustic overlap, the present study extended the investigation to focus on several factors that influence the amount of this overlap.

The acoustic analyses demonstrated that the post-release decay times of piano tones decrease as pitch (register) increases and as the duration of the tone increases. These effects are due in part to pre-release decay, which is faster for high than for low tones and more extensive for long than for short tones, and in part to an increase in post-release decay rate with pitch (but not with duration). This information, previously unavailable in systematic form in the literature, led to the prediction that listeners would adjust their perceptual criteria in judging *legato* articulation, and that pianists would adjust their *legato* playing, so as to avoid extensive acoustic overlap. That is, KOTs (the directly observable variable in MIDI recordings) were expected to be longer for tones with shorter post-release decay times (i.e., high and long tones).

These predictions were confirmed in the perceptual experiment, which was really a study of "passive" *legato* production, without involvement of the fingers. Listeners' adjustments were evidently sensitive to acoustic overlap, with KOTs being longer for high and long tones. Whether perceptual constancy in terms of some criterion of auditory (non)overlap was maintained across conditions could not be demonstrated directly, but it may be assumed that the subjects aimed for such a constancy, as this was essentially what the instructions requested them to do. It would be naive to expect a simple measure of acoustic overlap to correspond to such a perceptual constancy, but application of dynamic models of auditory processing may uncover an invariant auditory property in future research.

The results of "active" *legato* production were different. Although KOTs were longer for long than for short tones, this may be attributed to motoric factors. There was no effect of register; what seemed like one was probably due to a difference between hands. Thus pianists' playing seemed to reflect primarily motoric constraints, not adjustments to varying acoustic overlap. This implies, paradoxically, that the pianists' intended "best" *legato* might be judged by listeners (even by themselves!) to be nonoptimal in certain conditions, for example at a slow tempo in the low register (compare Figsures 5 and 4). A perceptual

evaluation of the recorded natural *legato* samples remains to be conducted. It should be noted that, unlike the stimuli in the perception experiment, the natural productions varied in timing and dynamics, and in KOT from one tone pair to the next.[21] These kinds of natural variation may well have an effect on the perceptual criterion for what constitutes a good *legato* in a melody.

There was one factor that both passive and active *legato* players were sensitive to: When the successive tones were relatively dissonant (1 st apart), adjusted and produced KOTs were shorter than when the tones were relatively consonant (3 st apart). Although, in principle, pitch separation *per se* could have been responsible for this difference, the effect was attributed to relative consonance on grounds of plausibility.[22] However, the effect may not be purely aesthetic in nature: Due to the larger number of shared harmonics of relatively consonant tones, there may be greater masking and/or fusion between consonant than between dissonant tones (see DeWitt & Crowder, 1987), making acoustic overlap more difficult to detect. There was also a confounding of step size with pitch range, and hence with average pitch and average acoustic overlap: On the average, *arpeggi* had slightly shorter overlaps than chromatic scales because they covered a one-octave range, whereas the latter extended only over the first 4 st in this range. More detailed investigations will be required to sort out these variables.

Palmer (1989) and Drake and Palmer (1993) reported that KOTs tended to be shorter when one of two successive tones was relatively long, especially the first one. This observation seems to contradict the clear tendency towards longer KOTs for longer tones in the present study. However, their finding was obtained in materials containing both short and long tones, and their pianists were not instructed to play *legato* throughout. Most likely, their finding reflects the fact that long tones tend to end motives and phrases; the reduced KOT following such tones is an aspect of "phrasing" which was absent in the present homogeneous materials, unless the progressive reduction in KOT during the ascending and descending portions of the scales and *arpeggi* is to be interpreted in a similar manner.

MacKenzie and Van Eerd (1990) observed longer KOTs for the right hand than for the left hand in rapid scale playing. The present production data, obtained at much slower tempi, tend to support their interpretation of this effect as one of hand, rather than of register. It appears that, in some pianists at least, the right hand is more adept at playing *legato* than the left hand.

The absolute KOTs reported here are longer than those obtained in most earlier studies. This is probably due to the relatively slow tempi and to the explicit instruction to play *legato*, though it is also possible that the Roland RD250s keyboard, with its relatively easy action, encouraged an increased *legato*. Certainly, replications and extensions of the present findings on a real computer-monitored piano are desirable, together with measurements of the post-release decay times of natural piano tones.

The relative insensitivity of the present pianists to acoustic/perceptual factors should not be interpreted as implying that pianists cannot adjust their degree of *legato* when musical circumstances require it. On the contrary, they are likely to be quite flexible in that regard. Kuwano et al. (1994) demonstrated that a pianist can play with different degrees of connectedness or separation when instructed to do so, though their results look somewhat categorical, with little difference in acoustic overlap between tones intended to be "marginally connected," "overlapping a little," and "overlapping to some degree." However, these terms are nearly synonymous and may have been so understood by the pianist. Whether there is a tendency to perceive or produce *legato* in a categorical fashion is an interesting question for future research, as are the many factors that may affect KOT, such as fingering, phrasing, and individual style. All expressive parameters of music performance are subject to continuous and systematic modulation, and KOT is hardly an exception. However, it is a little studied aspect of piano technique, and much remains to be learned about it.

REFERENCES

Benade, A. H. (1990). *Fundamentals of musical acoustics*. New York: Dover. (Originally published in 1976 by Oxford University Press.)

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Buus, S., & Florentine, M. (1985). Gap detection in normal and impaired listeners: The effect of level and frequency. In A. Michelsen (Ed.), *Time resolution in auditory systems* (pp. 159-179). Berlin: Springer-Verlag.

Collyer, C. E. (1974). The detection of a temporal gap between two disparate stimuli. *Perception and Psychophysics, 16*, 96-100.

Dannenbring, G. L., & Bregman, A. S. (1976). Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 544-555.

DeWitt, L. A., & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals: The oomph in Stumpf. *Perception and Psychophysics, 41*, 73-84.

Divenyi, P. L., & Danner, W. F. (1977). Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception and Psychophysics, 21,* 125-142.

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10,* 343-378.

Fitzgibbons, P. J., Pollatsek, A., & Thomas, I. B. (1974). Detection of temporal gaps within and between perceptual tonal groups. *Perception and Psychophysics, 16,* 522-528.

Formby, C., & Forrest, T. G. (1991). Detection of silent temporal gaps in sinusoidal markers. *Journal of the Acoustical Society of America, 89,* 830-837.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception and Psychophysics, 36,* 359-368.

Kameoka, A., & Kuriyagawa, M. (1969). Consonance theory part II: Consonance of complex tones and its calculation method. *Journal of the Acoustical Society of America, 45,* 1460-1469.

Kuwano, S., Namba, S., Yamasaki, T., & Nishiyama, K. (1994). Impression of smoothness of a sound stream in relation to *legato* in music performance. *Perception and Psychophysics, 55,* 173-182.

MacKenzie, C. L., & Van Eerd, D. L. (1990). Rhythmic precision in the performance of piano scales: Motor psychophysics and motor programming. In M. Jeannerod (Ed.), *Attention and Performance XIII* (pp. 375-408). Hillsdale, NJ: Erlbaum.

Martin, D. W. (1947). Decay rates of piano tones. *Journal of the Acoustical Society of America, 19,* 535-541.

Moore, B. C. J., & Glasberg, B. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America, 74,* 750-753.

Namba, S., Kuwano, S., & Yamasaki, T. (1992). Distinction between legato and staccato styles of rendition in piano playing: An illusion. presented at the Second International Conference on Music Perception and Cognition, Los Angeles, CA.

Neff, D. L., Jesteadt, W., & Brown, E. L. (1982). The relation between gap discrimination and auditory stream segregation. *Perception and Psychophysics, 31,* 493-501.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 331-346.

Perrott, D. R., & Williams, K. N. (1971). Auditory temporal resolution: Gap detection as a function of interpulse frequency disparity. *Psychonom. Sci.* 25, 73-74.

Plomp, R. (1964). Rate of decay of auditory sensation. *Journal of the Acoustical Society of America, 36,* 277-282.

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical band-width. *Journal of the Acoustical Society of America, 38,* 548-560.

Repp, B. H. (1993). Some empirical observations on sound level properties of recorded piano tones. *Journal of the Acoustical Society of America, 93,* 1136-1144.

Repp, B. H. (1994). Relational invariance of expressive microstructure across global tempo canges in music performance: An exploratory study. *Psychological Research, 56,* 269-284.

Repp, B. H. (in press). Pedal timing and tempo in expressive piano performance: A preliminary investigation. *Psychology of Music.*

Shailer, M. J., & Moore, B. C. J. (1983). Gap detection as a function of frequency, bandwidth, and level. *Journal of the Acoustical Society of America, 74,* 467-473.

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, 35A,* 377-396.

Sloboda, J. A. (1985). Expressive skill in two pianists: Metrical communication in real and simulated performances. *Canadian Journal of Psychology, 39,* 273-293.

Weinreich, G. (1990). The coupled motion of piano strings. In Five Lectures on the Acoustics of the Piano, edited by A. Askenfelt (Royal Swedish Academy of Music, Stockholm), pp. 73-81.

Williams, K. N., & Perrott, D. R. (1972). Temporal resolution of tonal pulses. *Journal of the Acoustical Society of America, 51,* 644-647.

Wogram, K. (1990). The strings and the soundboard. in Five Lectures on the Acoustics of the Piano , edited by A. Askenfelt (Royal Swedish Academy of Music, Stockholm), pp. 83-99.

FOOTNOTES

Journal of the Acoustical Society of America, 97, 3862-3874 (1995).

[1]Alternatively, the damper pedal may be engaged while depressing keys successively. The present study, however, was not concerned with pedaling.

[2]The technique of "partial release," which is possible on grand pianos, will not be considered here.

[3]Sonoko Kuwano (Namba, Kuwano, and Yamasaki, 1992) has pointed out that the overlap is distinctly audible as an "impurity" at the onset of an isolated tone excerpted from a legato passage. Note the parallel to the perception of coarticulation in speech (e.g., Fowler, 1984).

[4]The spectral composition of the sounds was not reported.

[5]Although they failed to mention this in their published paper, Kuwano et al. did include the original key release when playing back the individual tones (Kuwano, personal communication). Thus they had information about post-release decay times as well as about KOTs, but only acoustic tone overlap times were reported.

[6]Sloboda seemed to be unaware of the post-release decay of piano tones, as he stated that "This event [i.e., release of the key] is simultaneous with the termination of the sound" (p. 383).

[7]Extreme *legato* is also referred to as *legatissimo* or "finger pedaling." Nearly continuous use of the damper pedal in these performances further increased the acoustic overlap of tones and may also have increased the tolerance levels of listeners.

[8]On the Roland RD-250s, the spectral characteristics of the tones change with key velocity (intensity), as they do on a real instrument (cf. Repp, 1993). The choice of the fixed velocity value was arbitrary.

[9]The author is striving toward a consistent terminology of musical events. *Nominal tone duration* refers to the MIDI level of description and should be distinguished from *(actual, acoustic) tone duration*, which includes the post-release decay time, and from *note duration* (better: *note value*) which is not a physical magnitude at all but refers to the relative values (integer ratios) of notated symbols, such as 1/4 or 1/8.

[10]Since the onset of energy in the amplitude envelope started 15 ms (half the duration of the integration window) before the onset of energy in the waveform, these measurement points correspond to 235, 485, 735, and 985 ms, respectively, from energy onset in the waveform.

[11]The rise time was the time from the onset to the peak of the amplitude envelope, minus 30 ms. This estimate was reasonably accurate because of the fast rise and slow decay of the energy. A signal with zero rise time and no decay (i.e., a step function) would have a rise time of exactly 30 ms (the window duration) in the RMS amplitude envelope.

[12]In fact, preliminary analyses of acoustically recorded piano tones suggest that their decay characteristics are even more irregular than those of the synthetic tones analyzed here.

[13]These times, too, were adjusted for the window duration by subtracting 30 ms from the times measured in the envelope. Resolution of amplitude values was not fine enough to determine the -60 dB points, which had been used as the

criterion of tone offset by Kuwano et al. (1994). However, those time points can be extrapolated from the differences between the -20 dB and -40 dB time points (see also Figure 3), assuming that the post-release decay (in dB) was linear.

[14]Very similar results were obtained when, instead of taking the difference between the -20 and -40 dB time points relative to peak sound level, the time to decay by 20 dB was measured relative to the sound level at key release. Additional measurements showed that an increase in key velocity from 40 to 60 increased this decay time by about 10 ms regardless of pitch, but a further velocity increase from 60 to 80 had no effect. Key velocity varied in the *legato* production experiment, reported below.

[15]The intended IOIs were 250, 500, 750, and 1000 ms. However, when the sound output was later digitized and measured, it was discovered that the IOIs were consistently too long by 3.9%, due to an unrecognized problem with the Max software. This problem did not affect the accuracy of timing: The standard deviation of IOIs within the same sequence was less than 1 ms, probably representing just human measurement error. The nominal tone durations (key release times) were not affected either.

[16]Overshoot was possible (and did occur on a few trials) because the range of the slider was fixed; thus, fast tone sequences required an adjustment in the lower (left-hand) region of the slider, whereas slow sequences required an adjustment in the upper (right-hand) region.

[17]The melody used by Kuwano et al. contained both long and short tones. Presumably, they were referring to the length of the first tone in short-long and long-short sequences.

[18]This fingering would not be used for an extended chromatic scale, but it is quite natural for a 5-note chromatic scale, though many alternative fingerings are possible.

[19]This interpretation is very tentative, as there seems to be a step size effect for the left hand in the low register which is just as large as that for the right hand in the high register. However, neither the step size main effect nor the register by step size interaction was significant for the left hand.

[20]It would be interesting to observe the same pianists' KOTs in more natural playing, without explicit instructions to play *legato*. In fact, at the end of the experimental session four of the present subjects played a simple monophonic tune three times with their right hand in the middle register, to provide performance data for a different study. Three of them produced KOTs that were 40-50 ms shorter than in the most comparable experimental condition, but the fourth pianist produced times that were about 15 ms longer, on the average. Their "overlap profiles" for the tune were quite dissimilar, possibly due to different choices of fingering. Clearly, there is much more to be learned about the factors that cause systematic variation in KOTs.

[21]The average MIDI velocity in the pianists' playing was a good deal higher than the fixed velocity in the perception experiment, but this was partially offset by a relatively lower volume of the auditory feedback. As mentioned in an earlier footnote, however, the post-release decay characteristics of the digital piano tones did not vary much with dynamic level.

[22]It could be argued that the experiment should have been designed to separate these two factors from the outset. However, this is impossible without a radical change in fingering patterns and scale construction, which would raise new problems. For example, a pitch interval larger but more dissonant than the minor third (3 st) is the tritone (6 st), but it would permit only a 3-tone *arpeggio* with the fingering 1-3-5-3-1. Passing the thumb under the other fingers is undesirable in a study of *legato*, as it almost certainly reduces KOT.

Haskins Laboratories Status Report on Speech Research
1994, SR-117/118, 211-232

Pedal Timing and Tempo in Expressive Piano Performance: A Preliminary Investigation*

Bruno H. Repp

The timing of pedal depressions and releases was measured relative to key depressions and releases in two pianists' performances of Robert Schumann's "Träumerei" on an electronic instrument. Each pianist provided 9 complete performances, 3 at each of 3 tempi, which were analyzed by examining in detail those positions in the music in which the pedal was used consistently. The principal questions were whether and how pedal timing adjusts to changes in *global tempo* (across performances) and in *local tempo* (within performances): Do pedal release times, onset times, or change (onset minus release) times exhibit *absolute or relative invariance* across either or both of these tempo changes? The results do not suggest any simple answer, since neither type of invariance was observed consistently. Pedal timing emerges as having a complex pattern that is sensitive to local and global tempo changes in varying degrees, yet exhibits consistency across repeated performances by the same pianist. There were striking differences in pedal timing between the two pianists, who differed in level of skill.

INTRODUCTION

Modern pianos have two or three pedals, the most important of which is the *damper pedal* , referred to simply as "pedal" here. When depressed, it raises all dampers, so that the strings vibrating at that time continue to vibrate until the pedal is raised or until the vibrations decay naturally. It also enables other strings to vibrate sympathetically, thus enriching the timbre of the sustained sounds. Once the basic skill of hand-foot coordination has been mastered by a piano student, pedaling becomes a subconscious, automatic activity for most players. Concert pianists and teachers naturally examine and refine their pedaling skills using auditory feedback to guide them, and great artists often exhibit a masterful pedaling technique which contributes to their characteristic "tone" and "touch." Interestingly, as Heinlein (1929a) has observed, listeners tend to attribute the sonic consequences of pedaling to the pianist's manual skill, being unaware of how crucially piano performance depends on what the right foot is doing.

Piano scores often indicate when the pedal should be depressed and when it should be released. Just as often, composers and editors omit pedaling instructions from the score or insert them only at crucial points. Typically, a pianist pedals much more frequently than is indicated in the score. One common use of the pedal is to create smooth transitions between tones or chords that are difficult or impossible to connect by fingering alone. These transitional uses of the pedal (also referred to as syncopated or *legato* pedaling) are rarely notated and are at the pianist's discretion. The present study focuses on this type of pedaling.

For the psychologist interested in the cognitive and kinematic processes involved in music performance, pedaling raises interesting questions about motor control, coordination, and the role of auditory feedback, and it offers an opportunity to study an important component of pianistic skill about which rather little is known from a scientific perspective. Qualitative and quantitative aspects may be distinguished: *pedal use* and *pedal timing*, respectively. Musical

This research was supported by NIH grant MH-51230. I am grateful to LPH for lending her artistry to this project, and to Nigel Nettheim, Caroline Palmer, and Henry Shaffer for helpful comments on an earlier version of the manuscript.

notation conveys only instructions about pedal use, if any, and observation of pianists (including self-observation) with the naked eye and ear similarly yields only qualitative information. Empirical questions about pedal use concern the relative frequency with which pianists depress the pedal in a given piece of music, where in the music they use it (and why), and when they depress and release the pedal relative to the notated musical events (described in qualitative terms such as before, after, and between). In contrast, the precise timing of pedal actions is an aspect of the *expressive microstructure* of music performance (Clynes, 1983), whose measurement requires special instrumentation. Questions about pedal timing concern the detailed temporal relations between hand and foot actions, as measured by the exact times elapsing between key and pedal depressions, as well as the timing of successive foot actions: Are these intervals invariant or context-dependent? Do they stretch and shrink with changes in tempo? Do pianists differ in their pedal timing characteristics? Can pedaling skill be measured objectively? The basic technology to address these questions has been available for some time (e.g., Heinlein, 1929b; Seashore, 1938), but pedal timing has been little investigated. The recent proliferation of MIDI technology, however, greatly facilitates such studies.

The only systematic studies of pedaling in the psychological literature known to this author were conducted by Heinlein (1929b, 1930) and recently by Taguti, Ohgushi, and Sueoka (1994). Heinlein (1929b) compared pianists' pedaling patterns in a qualitative way, by counting the number of times the pedal was depressed and by examining rough graphs (kymograms) of pedal actions relative to the onsets of the musical tones. He pointed out large differences in pedal use among different pianists playing the same music and considerable variability in pedaling even in the same pianists' repeated performances of the same music (cf. also Banowetz, 1985: p. 9). Heinlein (1930) asked four pianists to pedal while playing, tapping, listening to, or imagining the same music. He found that it is virtually impossible to produce a good pedaling pattern without actually playing the music. This confirms a point made in many discussions of pedaling in the pedagogical literature, namely that it is "governed by the ear" (see, e.g., Marek, 1972; Neuhaus, 1973; Newman, 1984; Philipp, 1984; Banowetz, 1985). The recent study by Taguti et al. (1994), based on 8 pianists' performances of a Chopin Waltz in three expressive styles, did not focus on the pedaling patterns as such but rather on the multidimensional structure of their dissimilarities and its relationship to verbal descriptions of performance quality. At this time, no detailed study of pedal timing has been reported in the literature.

To explain how pedal timing was measured in the present study, Figure 1 schematically illustrates a typical *legato* pedaling pattern in terms of MIDI events. Two successive *legato* melody notes are shown symbolically, and a relatively slow tempo is assumed, allowing the player to pedal with each individual note if (s)he so wishes.

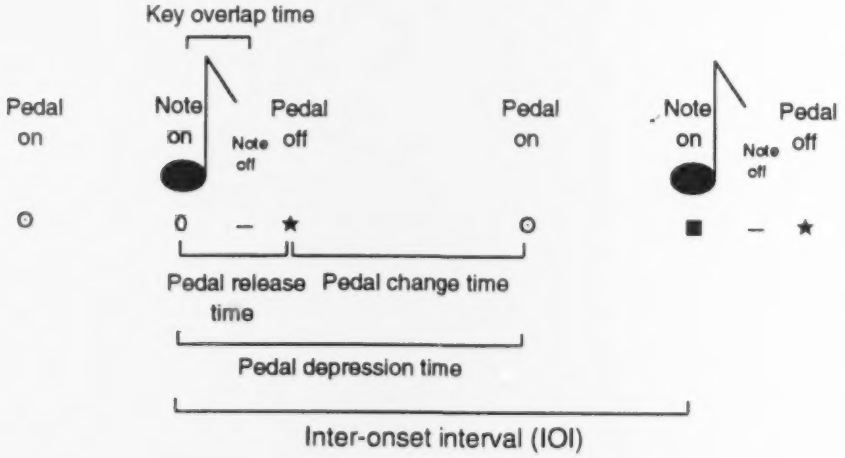


Figure 1. Schematic illustration of MIDI events and temporal intervals in a typical *legato* pedaling pattern. The symbols are the ones that will be used in later data graphs (Figs. 6-7). The first "note on" corresponds to time zero.

The "note on" events (which closely coincide with the acoustic onsets of the tones represented by the notes) define a physical *interonset interval (IOI)* within which pedal events may be located. In *legato* pedaling, a pedal depression ("pedal on") typically precedes a key depression ("note on"), and a pedal release ("pedal off") follows it. Thus the pedaling serves to prolong the duration of the *preceding* tone whose key release ("note off") may occur before or after the onset of the following tone, but nearly always before the pedal is released; thus the *legato* connection of the tones is enhanced by the prolongation afforded by the pedaling. (If the key release *follows* the pedal release, the pedaling is redundant with regard to the achievement of *legato* articulation but adds richness of timbre.) This pedal action is repeated for the next tone, and so on. Within an IOI defined by two key depressions, then, there are typically two pedal events: a release followed by a depression. This rapid sequence of foot actions is known as a *pedal change*.

Three time intervals will be of particular interest in this study (see Figure 1). The first is the temporal location of a pedal release within an IOI, referred to as *pedal release time (PRT)*. The second is the location of a pedal depression within an IOI, referred to as *pedal depression time (PDT)*. The third is the interval between a pedal release and a pedal depression, provided that both occur within (or very nearly within) the same IOI; it will be referred to as *pedal change time (PCT)*. Each of these three intervals can be specified in *absolute* terms, in milliseconds, or in *relative* terms, as a percentage of the IOI in which it occurs (PRT%, PDT%, PCT%).

Because pedaling depends on what the hands are doing, it seems that it must be *rhythmically coordinated* in some fashion with the manual actions. The nature of this coordination would be most easily investigated when all successive IOIs are equal in duration, as might be the case in simple exercises carried out in a mechanical fashion. The present study, however, focuses on the more complex but also more ecologically valid situation of expressive music performance, in which IOI duration is highly variable (see, e.g., Repp, 1992). The duration of a given IOI depends on three factors: (1) the note value specified in the score, (2) the *global tempo*, and (3) *local tempo* (expressive timing, agogic variation). The first factor played no role in the present study, as all IOIs examined corresponded to eighth notes (half beats) in the score. Attention thus focused on the second and third factors. Global tempo accounts for systematic differences in the duration of the same IOI across performances played at different tempi; local tempo accounts for systematic differences among different IOIs within the same performance. Variations in local tempo are largely governed by structural factors (hierarchical grouping, melodic contour, harmony, etc.), whereas global tempo is structure independent. Given that IOIs vary in duration due to these two sources, the theoretical question addressed in this study was *how pedal timing adjusts to these variations*.

The relative precision of pedal timing is unknown at present. It could be that pedal timing (like pedal use) is highly variable, unlike tone onset timing, which is controlled very precisely and is highly replicable across repeated performances of the same music (see, e.g., Seashore, 1938; Shaffer, 1981; Shaffer, Clarke, & Todd, 1985; Repp, 1992). Replicability of pedal timing patterns across repeated performances at the same tempo is a prerequisite for the investigation of systematic adjustments to tempo changes. Such adjustments may take either (or neither) of two forms: *absolute invariance* or *relative (relational) invariance*. Absolute invariance would hold if pedal releases and/or depressions always followed tone onsets by a fixed interval, or if pedal change times were approximately constant, regardless of the duration of the IOI within which these events are situated. The possible absolute invariance of pedal change times in particular seemed an interesting hypothesis, given the ubiquity, rapidity, and automaticity of the release-depression action sequence. These absolute invariance hypotheses may be contrasted with the corresponding relative invariance hypotheses, according to which some or all of the intervals mentioned stretch and shrink proportionally with changes in IOI duration.[1] Proportional changes in timing microstructure with changes in global tempo have been observed in many other skilled motor behaviors, at least to a first approximation (see Gentner, 1987; Heuer, 1991). Finally, it is possible that neither of these simple hypotheses applies, and that adjustments of pedal timing to changes in manual timing are more complex, but nevertheless systematic. It is also conceivable that a given temporal interval (pedal release time, say) is absolutely invariant across global tempo changes but relationally invariant across local tempo changes.

The hypothesis of absolute invariance has some plausibility because the acoustic decay characteristics of piano tones do not change with

tempo. If the purpose of pedaling is to "catch" a tone at a certain dynamic level and prolong it, pedal depression times may well be insensitive to changes in tempo. However, a pianist must also avoid depressing the pedal when tones that should not be prolonged are still sounding. For that reason, pedaling is likely to be sensitive not only to the timing of key depressions (which served as temporal reference points in this study) but also to that of key releases. Piano tones do not cease immediately following key release but decay over a few hundreds of milliseconds, with low tones decaying more slowly than high tones (Repp, 1995). This decay places a constraint on pedal depression times, and if key release times vary with tempo, so may pedal depression times.

The time between the key depression for one tone and the key release for the preceding tone will be referred to as *key overlap time (KOT);* it is positive when there is overlap (as in Figure 1) and negative when there is a gap. Gaps are usually bridged by pedaling and thus are inaudible. They may occur when it is difficult to connect two tones with the fingers, but also in other places, as finger *legato* is not strictly necessary when there is *legato* pedaling. However, when finger *legato* is possible, it is commonly maintained even when the pedal is being used; the resulting key overlap time can be considerable and may vary with structural factors. For example, two consonant tones may be overlapped more than two dissonant tones, and the following pedal change may be correspondingly delayed. Thus, factors influencing overlap time may also influence pedal timing. For this reason, some key overlap times were also measured in the present study.

Finally, individual differences in pedal timing were of interest, as previous studies (Heinlein, 1929b, 1930) had focused only on individual differences in pedal use. Differences in pedal timing among pianists may reflect differences in motor organization responsible for differences in "tone" and "touch," as well as differences in level of technical skill.

These various issues were examined in MIDI data obtained from 18 integral performances of Robert Schumann's well-known piano piece, "Träumerei" (op. 15, No. 7), which have been the subject of two previous studies by this author (Repp, 1994a, 1994b). They derive from two pianists, each of whom played the piece three times at each of three different global tempi. Repp (1994a) addressed the question of whether expressive timing patterns (i.e., IOI durations) expand and contract proportionally with changes in global tempo (i.e., whether they show relational invariance) or whether they change in a more complex way. (Of course, absolute invariance is impossible in this case.) Some small but statistically significant deviations from relational invariance were noted, and Desain and Honing (1994) have reported larger deviations from relational invariance in a different piece of music. For the present purposes, it is sufficient to note that variations in global and local tempo were reasonably independent in the performances studied, as was re-confirmed in the statistical analyses reported below.

Although Repp's (1994a) study focused primarily on tone interonset intervals, it also included a selective analysis of pedal timing, restricted to 8 recurrences of one particular IOI in the musical structure which always contained a pedal change (bar 1-1 and corresponding locations, which represents a quarter-note IOI; see Figure 2 below). These data were not very clear with regard to the two invariance hypotheses, but they revealed striking individual differences between the two pianists, both in pedal timing and in its sensitivity to tempo variation. The purpose of the present study was to analyze the pedal timing data from these performances in more detail, so as to examine more thoroughly the influence of tempo changes on the timing of hand-foot coordination in expressive piano performance. The present study was *not* concerned with providing a detailed explanation of pedal use and timing as such, i.e., with accounting for why and how the pedal was deployed at particular points in the music, though some pertinent comments will be made. The primary focus was on the sensitivity of the pedal timing pattern (whatever it happened to be) to changes of tempo.

Each pianist's data were analyzed separately. Based on an initial analysis of pedal use, sets of structurally similar points in the music ("vertical slices," cf. Figure 2 below) were selected where the pedal was changed consistently. Six such sets were analyzed, capturing about half of all pedaling events. By conducting statistical analyses within structurally relatively homogeneous sets, variation in local tempo was to some extent dissociated from structural variation in the music, though even within-set variation in local tempo, to the extent that it was not random or idiosyncratic, presumably was still determined by structural features of the music. Within each analysis set, the effects of global tempo and of local tempo, as well as their interaction were assessed for each of the several temporal intervals

of interest. In all these repeated-measures ANOVAs, the error term for each effect of interest was its variation across performances with *the same global tempo*. In that way, across-performance stability provided the criterion for assessing the statistical reliability of any effect. Analyses were performed both on absolute and relative measures of pedal timing (PRT, PDT, PCT; PRT%, PDT%, PCT%): If an absolute interval does not vary with (global or local) tempo but the corresponding percentage does, the absolute invariance hypothesis is supported. If the opposite result is found, the relational invariance hypothesis is supported. If neither measure shows a significant effect, the results are inconclusive; high variability may be to blame. If both show a significant effect, some more complex type of tempo adjustment is suggested. No statistical comparisons were conducted between the six analysis sets, which represented different positions in the musical structure.

Method

The music

The score of Schumann's "Träumerei" is shown in Figure 2, laid out on the page so that structurally similar parts are aligned vertically. Since the first 8 bars are repeated, the music comprises 24 measures. There are three 8-bar sections (the general form is A-B-A'), each of which contains two 4-bar phrases. The predominant note value is the eighth note; thus most IOIs are nominally equal, though their actual durations varied dramatically, due to expressive timing (see Repp, 1994a). For a more detailed discussion of the music and its structure, see Repp (1992).

Pianists

Two pianists provided the performances: LPH, a professional musician in her mid-thirties, and BHR, the author, a serious amateur in his late forties. Both pianists were thoroughly familiar with the music and had played it many times before. Although BHR was capable of playing the piece accurately, consistently, and with good expression (cf. Repp, 1994a), his technical skills were clearly much less developed than LPH's; this was expected to be reflected in the pedal timing data.

Recording procedure

The recording procedure is described in detail in Repp (1994a, 1994b). The instrument was a Roland RD-250s digital piano with weighted keys and DP-2 foot pedal switch. Although this simple pedaling device did not permit degrees of pedal depression and did not simulate sympathetic string vibration, it was nevertheless believed that the pianists' habitual pedaling patterns would be transferred to the electronic instrument, perhaps with some automatic adjustments to its acoustic characteristics (as would also occur with any unfamiliar acoustic piano). The digital piano was connected to a microcomputer which stored the performances in MIDI format (note on and off times, key velocities, and pedal on and off times), with a temporal resolution of 5 ms. The pianists monitored the sound ("Piano 1") over earphones. Each pianist played the piece 3 times at each of 3 aesthetically acceptable tempi, called "slow," "medium," and "fast" in the following. Each tempo was cued by a metronome, which was turned off before the performance started. (See Repp, 1994b, for the exact metronome settings and observations on the pianists' relative accuracy in following them.) The performances within each tempo category naturally differed somewhat in global tempo, but those differences were small relative to the differences between tempo categories.

Analysis

Following an initial analysis of qualitative pedal usage, 6 "vertical slices" were taken through the score, each yielding 8 structurally identical or similar positions, 4 bars apart and referred to by bar and eighth-note numbers (e.g., 13-1 denotes the first eighth note in bar 13). In computing IOIs, whenever several tone onsets coincided nominally but their exact onset times were not identical (as is usually the case), the onset time of the tone with the highest pitch was taken as the reference. Pedal release and depression times within these IOIs were measured, and if there were two pedal events within the IOI, typically a release followed by a depression, the pedal change time was also calculated. Furthermore, the key release time of the preceding melody tone was determined if it fell within or close to the onset of the IOI. All these temporal measures were expressed both in milliseconds and as percentages of the total IOI. Within each analysis set (vertical slice through the score), each of these absolute and relative values was subjected to a mixed-model ANOVA, separately for each pianist, with the fixed factors of (global) tempo (3 levels) and position (i.e., local tempo; 8 levels, or less when there were missing data), and performances (3 levels, nested within tempo categories) as the random factor. Because a very large number of F values was computed, they will not be reported in detail. The statistical results are summarized in tabular form after a descriptive presentation of selected data.

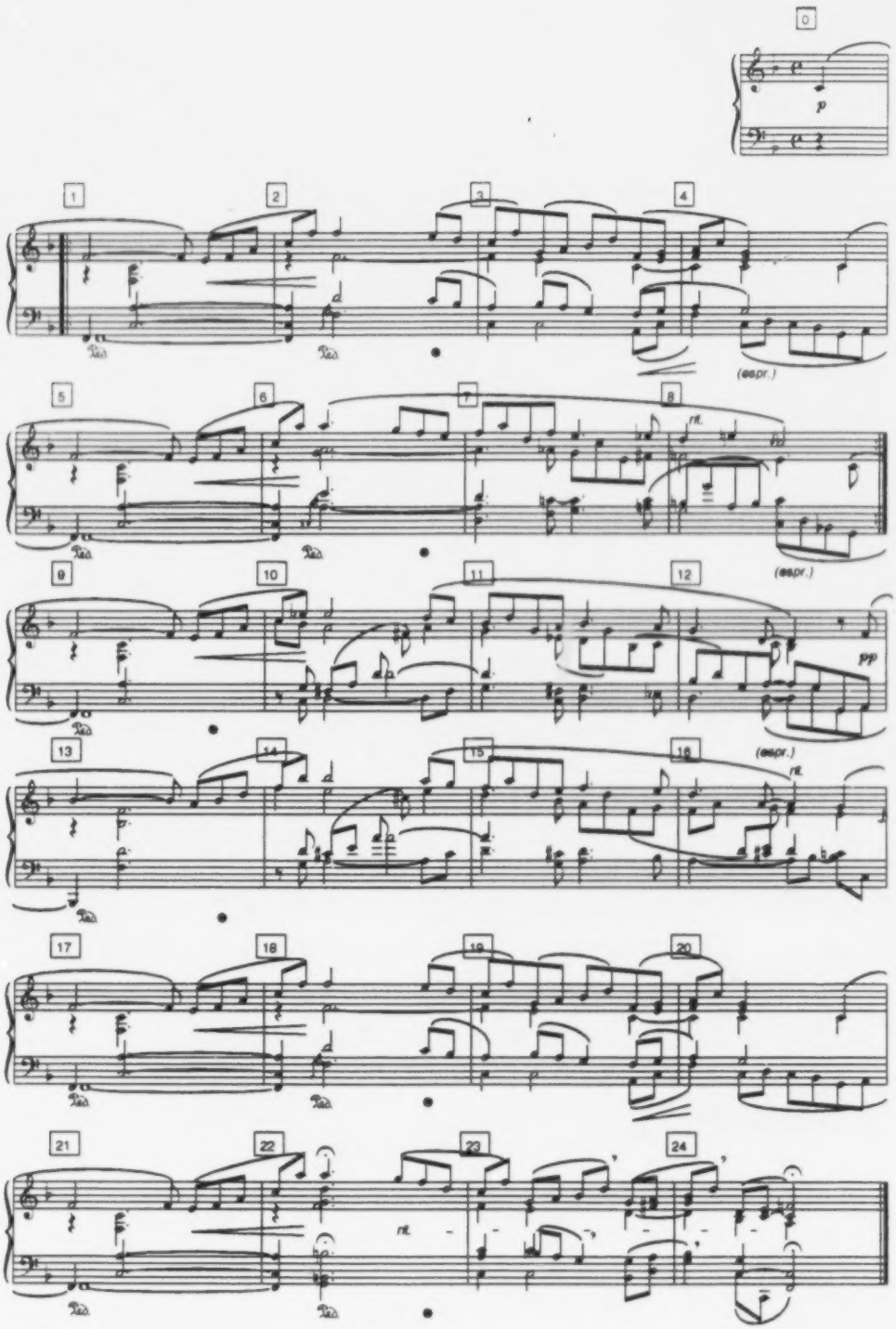



Figure 2. **Score of Robert Schumann's "Träumerei" (No. 7 from "Kinderszenen," op. 15). The computer-generated score follows the Clara Schumann edition (Breitkopf & Härtel), except for some minor deviations due to software limitations. The page layout helps clarify the musical structure.**

Results and Discussion

Pedal use

Table 1 shows the total frequencies of pedal use in the 9 performances by each pianist. Each frequency represents a pair of events: pedal depression followed by pedal release (regardless of the interval in between). It is evident that LPH used the pedal more often than did BHR. Both pianists show a tendency to use the pedal less frequently as the tempo increased. However, due to variability within each tempo category, this tendency was not significant. Heinlein (1929b) compared two famous pianists' highly divergent pedaling in the same music; their total frequencies were 51 and 135, respectively. The present counts fall between these two extremes.

Table 1. *Frequencies of pedal use in the 18 individual performances.*

Tempo	Perf	LPH	BHR
Slow	1	88	74
	2	93	73
	3	91	69
Medium	1	98	75
	2	87	68
	3	82	75
Fast	1	85	67
	2	92	67
	3	81	66

The detailed distribution of pedal usage throughout the music is shown in Figures 3 (LPH) and 4 (BHR). The layout of these figures matches that of the score in Figure 2. The frequencies plotted are summed over the 9 performances of each pianist; for bars 1-8, moreover, the data have also been summed over the within-performance repeat, so that there were 18 renditions altogether. The white and black bars represent the frequency of pedal releases and depressions, respectively, within the IOI starting on the half-beat indicated on the abscissa. The quarter-note IOI associated with the initial upbeat of the piece (bar 0), which usually contained the first pedal depression, is omitted in these figures. (The initial depression frequencies are the complement of the release frequencies in bar 1-1.) It should be noted that the temporal order of pedal releases and depressions *within* each IOI is not represented in these figures. In the vast majority of cases, they followed the pedal change (off-on) pattern (as suggested by the relative placement of the bars in Figures 3 and 4), but there were instances of depressions preceding releases within IOIs in BHR's data.

A number of things can be observed in these figures. First, it is clear that the pedal was used much more frequently than prescribed in the score (Figure 2). Second, each pianist showed places where (s)he used the pedal in all performances (i.e., where the bars in the figure reach maximum height), whereas in other places the pedal was used less consistently. The two pianists' within-performance consistency may be gauged by comparing bars 1-4 (panel 1) with bars 17-20 (panel 5), which represent the identical music, and bars 9-12 (panel 3) with bars 13-16 (panel 4), which are very nearly transpositions of each other. BHR was somewhat more consistent than LPH by that comparison. Third, while there are similarities in pedal use between the two pianists, there are also many differences. The most striking difference is that LPH used the pedal change pattern all the time, so that the pedal was always down when a key was struck, whereas BHR showed some gaps in pedal use (see, e.g., the intervals 3-3 to 3-5 or 12-1 to 12-3) and also had a tendency to lift the pedal just before the next key depression, so that pedal changes sometimes straddled tone onsets. This may be a reflection of poor pedaling technique. Finally, it is evident that the pedaling patterns are relatively simple and sparse during the first half of each 4-bar phrase, but considerably more complex during the second half, which corresponds to changes in the melodic and harmonic complexity of the music (cf. Figure 2).

Pedal timing

Pedal change time distributions. Unlike pedal release times (PRTs) and pedal depression times (PDTs), whose determination was laborious, pedal change times (PCTs) could easily be obtained from the raw data by looking at pedal events only. Therefore, an initial rough analysis examined the distributions of PCTs across all performances within each tempo category. If pedal change is a stereotypical, reflex-like action pattern, the distribution of PCTs should exhibit a pronounced peak at some short duration, corresponding to the time needed to move the foot up and down. If there is absolute invariance, this peak should be independent of tempo. If the peak shifted with tempo, relational invariance would be indicated. This analysis included all PCTs, not just the ones in the 6 analysis sets.

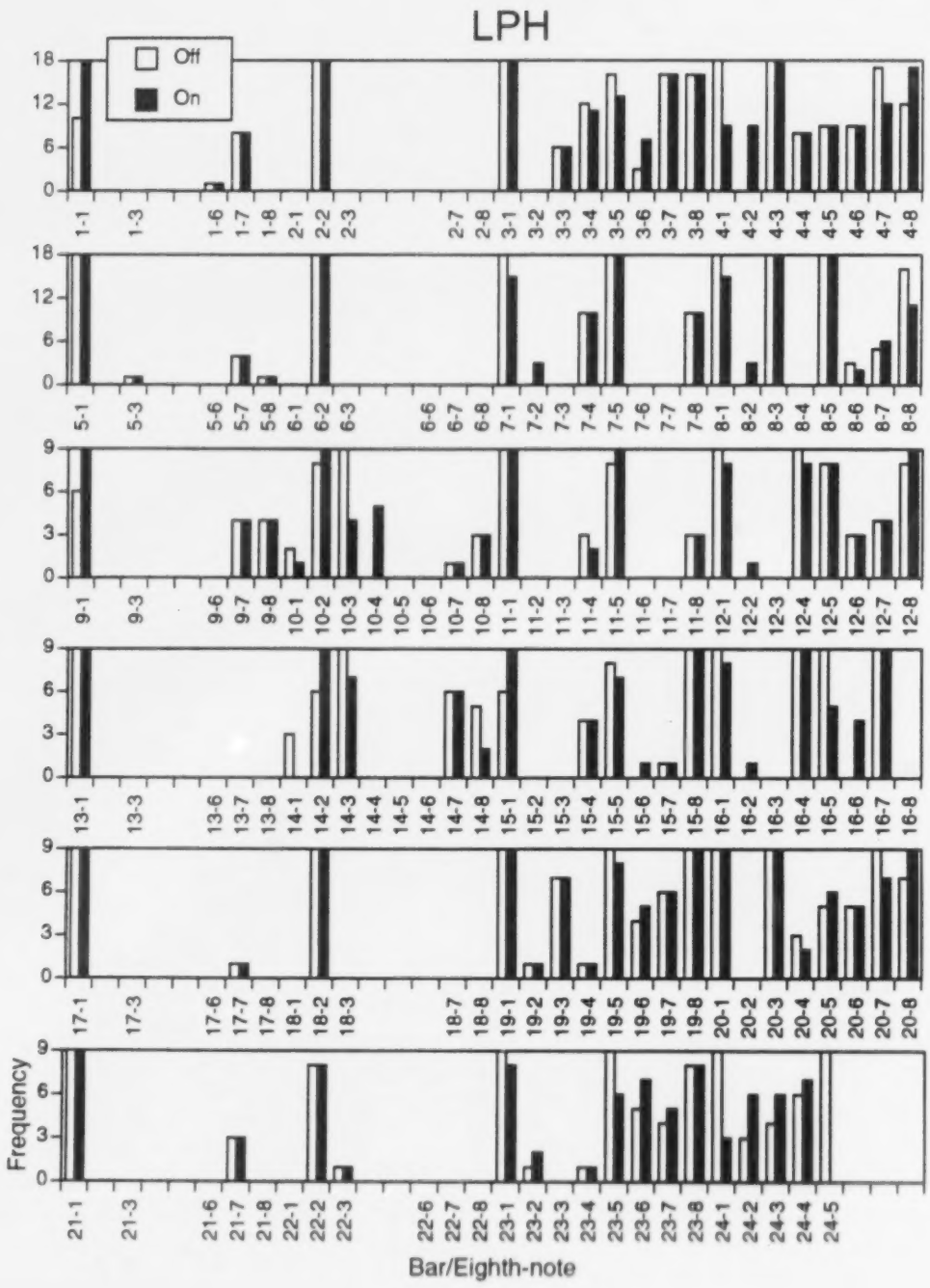


Figure 3. **Pedal use frequencies for pianist LPH. The layout of the figure corresponds to that of Figure 2, with the initial upbeat omitted. Each bar represents the frequency of pedal events within a particular IOI, added up over the 9 performances and, in the case of bars 1-8, over the two renditions within each performance.**

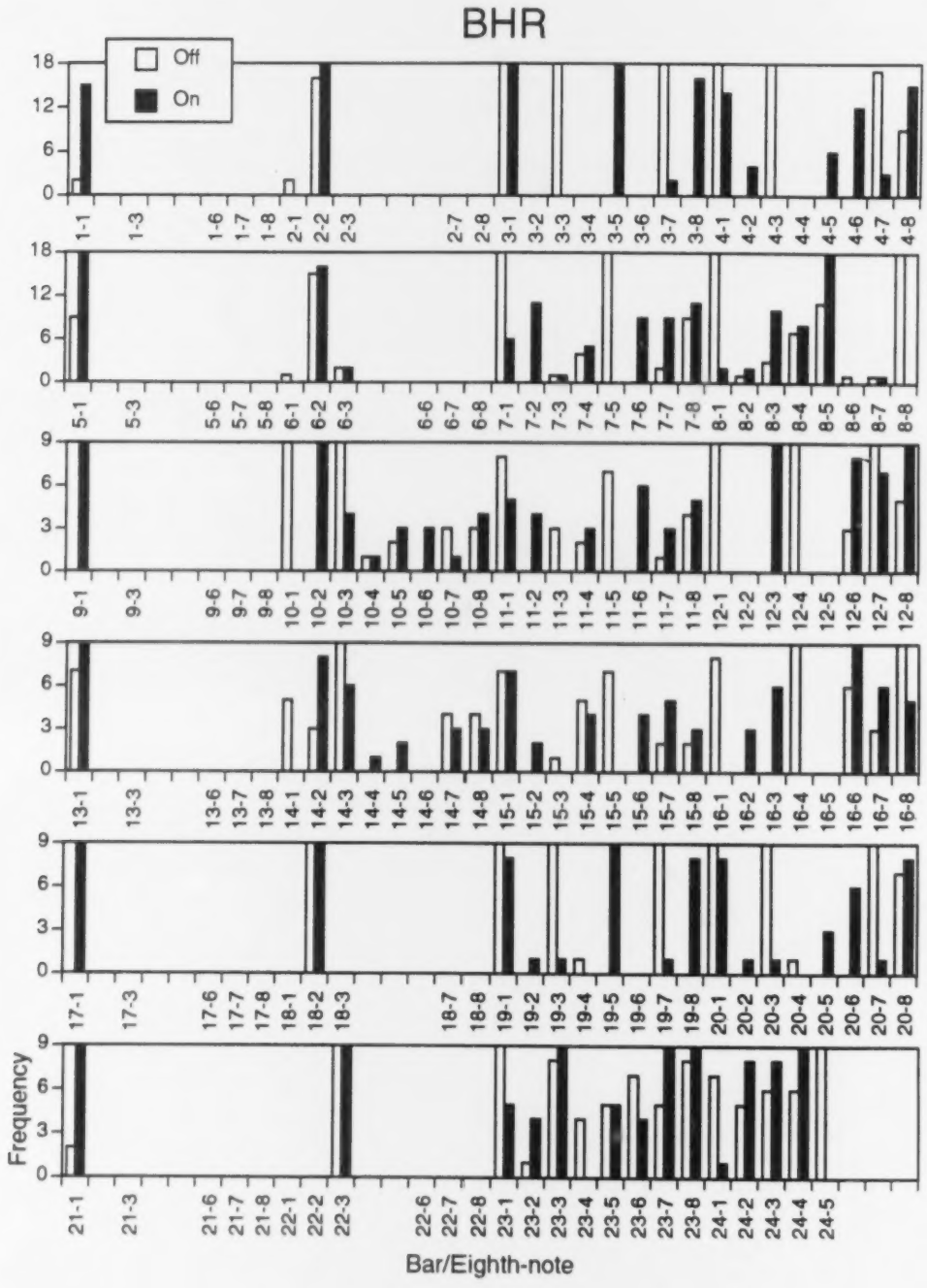


Figure 4. **Pedal use frequencies for pianist BHR.**

BEST COPY AVAILABLE

The PCTs across the 3 performances within each tempo category were sorted into 50-ms bins for each pianist. Since only short times were of interest, an upper limit of 500 ms was adopted. The resulting distributions are shown as line histograms in Figure 5. It can be seen that neither pianist exhibited a narrow peak. PCTs were broadly distributed between 50 and 400 ms. LPH did show a peak around 100-150 ms, but it was not narrow enough to suggest a fixed action pattern. BHR, who yielded fewer short PCTs, did not show any clear peak at all. Neither pianist's distributions suggest any shift with tempo.

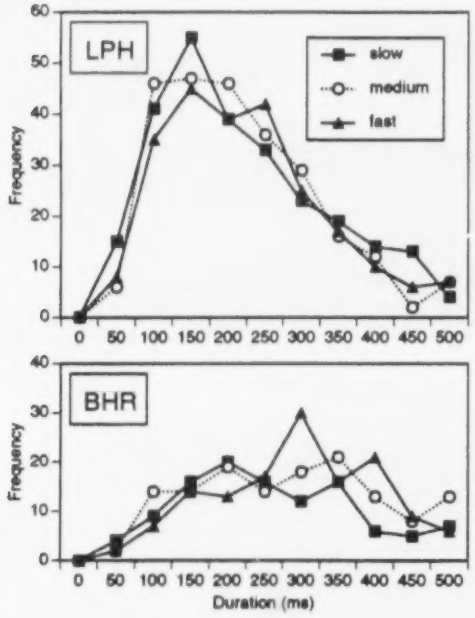


Figure 5. **Line histograms of pedal change times shorter than 500 ms in each pianist's performances at each of three tempi. Each data point represents a frequency count within a 50 ms bin whose *upper* bound is shown on the abscissa.**

This leaves open the possibility that PCTs, although they are obviously not absolutely invariant with respect to local tempo, are absolutely invariant with respect to global tempo. The following analyses examined this possibility and others at six structurally similar points in the score, which are aligned vertically in Figures 2-4. Because of the complexity of these analyses, however, only two of them will be presented in graphic detail.

Bar 2-2 and corresponding positions. This is the IOI that precedes the apex of each phrase. As the last IOI preceding a half-phrase boundary, it exhibits considerable expressive lengthening. It is special in two additional ways: It is the only eighth-note IOI bracketed (in the soprano voice) by tones of the same pitch, and it is the only IOI that contains two additional tone onsets within itself: In 5 of its 8 occurrences (bars 2-2, 6-2, 2-2R, 6-2R, 18-2), two grace notes (part of an arpeggiated chord in the left hand) occur during the IOI; these notes are identical in bars 2-2(R) and 18-2 but different in bar 6-2(R). There are also explicit pedal instructions in the score at these points. The grace notes are absent in bars 10-2 and 14-2, where the beginning of an imitative melodic motive appears in the tenor voice (left hand). Bar 22-2 also lacks the grace notes, and the chord is not arpeggiated, though LPH chose to play it in this fashion, which elongated the IOI (measured to the onset of the last and highest tone of the chord) enormously. For this reason, and also because BHR did not pedal at all during this IOI, bar 22-2 was excluded.

Figure 6 shows the timing data. Time runs from bottom to top here. The beginning of the IOI (the onset of the first tone) is on the abscissa, whereas its end (the onset of the second tone) is signified by a filled square. Each data point represents an average across the three performances at each tempo. Each compartment of the figure shows the data for the three global tempi, for one particular position in the music. Thus, effects of global tempo can be seen within compartments, effects of local tempo (position) across compartments.

IOI duration obviously decreased as global tempo increased. IOI duration also varied across the five IOIs containing grace notes, being generally longer in bar 6-2(R) than in bar 2-2(R), and longer in the repeats than in the first rendition of these bars (bar 18-2 being effectively a second repeat of bar 2-2). These differences represent variations in local tempo across structurally identical or highly similar positions in the music.

Both pedal releases and depressions occurred earlier in the two IOIs without grace notes, bars 10-2 and 14-2, than in the other bars, and somewhat earlier for BHR than for LPH. LPH's data for the "graceless" bars were not very informative, as not a single tempo or position effect reached statistical significance, for either absolute or relative measures of pedal timing. BHR, too, did not show any significant effects involving global tempo in those positions. However, his PRTs and PDTs occurred significantly earlier in bar 14-2 than in bar 10-2, both in absolute and in relative time, which contradicts any form of local invariance. Absolute or relative PCTs did not vary between those two positions.

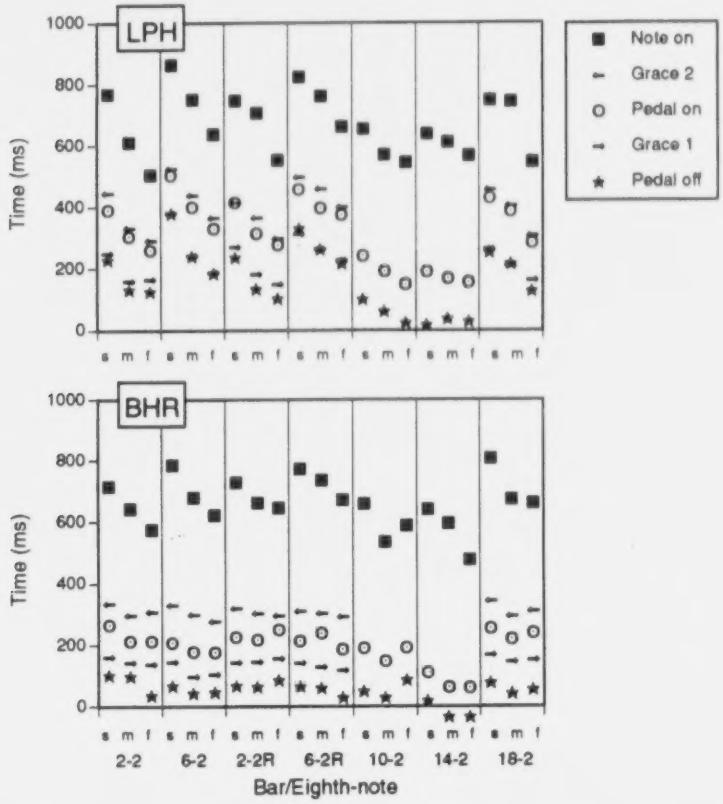


Figure 6. **Pedal timing data for bar 2-2 and corresponding positions. Each data point represents the average of measurements from three performances at each of three tempi: slow (s), medium (m), and fast (f). "Grace 1" and "Grace 2" represent grace note onsets. "Pedal off" marks pedal release time (PRT), "pedal on" marks pedal depression time (PDT), the interval between them is pedal change time (PCT), and "note on" marks the IOI duration.**

Figure 6 suggests that, in the IOIs containing grace notes, pedaling times and grace note onset times were coordinated. Consider first LPH's data. Her pedal releases immediately preceded, or coincided with, the onset of the first grace note, and pedal depressions immediately preceded the onset of the second grace note. All these times clearly varied with global tempo, and they also varied across positions in the music, occurring later in bar 6-2(R) than in bar 2-2(R). The tempo and position main effects were significant for all absolute measures, except for PCTs, which did not show a tempo effect. The time between grace note onsets, however, even though it seemed very similar to PCTs, did decrease somewhat with increasing tempo. By contrast, tempo effects were absent for most relative time measures, except for PRT% and PCT%. These data, then, provide evidence for relational invariance of grace note onset times (cf. Repp, 1994a), as well as of PDTs; paradoxically, however, they suggest absolute invariance of PCTs, even though these intervals seemed to be very similar to the intervals between grace note onsets. Differences across positions were also obtained for most relative time measures, except for PDT% and the interval between grace note onsets. PCT% showed a very strong position effect, being relatively shorter in bar 6-2(R) than in bars 2-2(R) and 18-2, whereas the interval between grace note onsets was relationally invariant across positions. Thus, pedaling times and grace note onsets did not behave quite in the same way, contrary to what one might conclude from a superficial inspection of the data.

BHR's results were quite different from LPH's, even though his pedaling times and grace note onsets also seemed coordinated. Here, however, they alternated rather than nearly coincided. Tempo effects were virtually absent here, both for

absolute and for relative measures. Significant tempo effects were observed only for the absolute and relative onset of the second grace note, which varied slightly with tempo but not enough to be relationally invariant. Most timing measures varied significantly across positions, however, and these differences tended to be larger for relative than for absolute measures. In contrast to LPH's data, most events occurred earlier in bar 6-2(R) than in bars 2-2(R) and 18-2, except for the absolute onset of the second grace note and both absolute and relative PCTs.

In summary, LPH's data suggest relational invariance of pedal depressions (along with grace note timing), but absolute invariance of PCTs. BHR's data are less clear but not inconsistent with a similar interpretation. Although pedal timing appears to be coordinated with grace note onsets, there is also evidence that it is not completely tied to those events. This was the only opportunity in these data to observe pedaling during relatively fast manual actions (grace notes) within an eighth-note IOI.[2]

Bar 3-1 and corresponding positions. The next vertical slice through the music was taken at bar 3-1 and the corresponding locations in bars 7-1, 3-1R, 7-1R, 11-1, 15-1, 19-1, and 23-1. In each case, the IOI starts with a chord, which accompanies the melodic line in the soprano voice; the end of the IOI is marked by the next, unaccompanied tone in the soprano voice (Figure 2). Figure 7 shows the quantitative results, which also include the key release ("note off") times of the preceding tone in the soprano voice that defines the key overlap time (KOT).

Consider first LPH. Naturally, her IOI durations varied with tempo, but they also varied across positions: The IOI was shorter in the middle section (bars 11-1 and 15-1) and longer near the end of the piece (bar 23-1). Key releases of preceding tones occurred extremely late, reflecting a *legatissimo* playing style. KOTs varied significantly with tempo and especially across positions. PRTs, too, varied with tempo and across positions, and they usually fell in the vicinity of the key release times for the preceding tone.

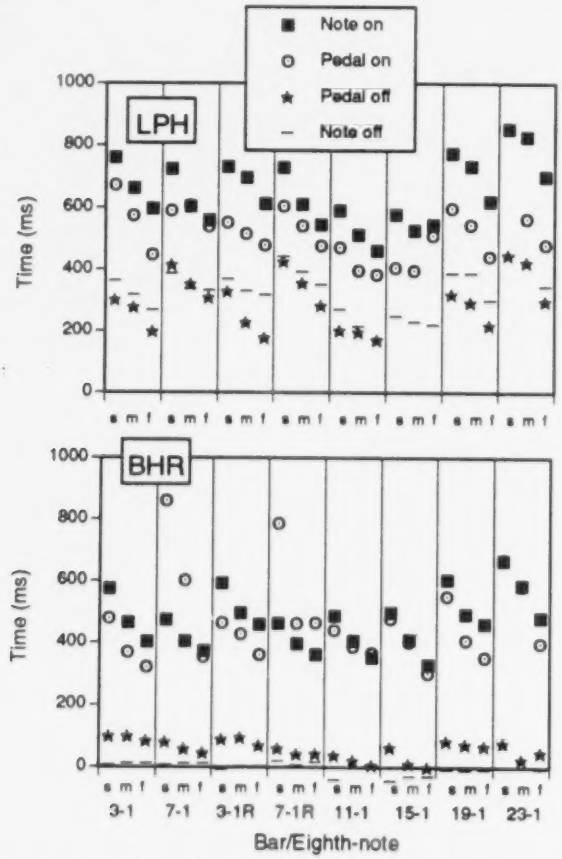


Figure 7. **Pedal timing data for bar 3-1 and corresponding positions. "Note off" marks key overlap time (KOT).**

(PRTs were highly variable in bar 15-1, so these data are not shown in the figure and were not included in the statistical analysis.) While LPH's PRTs clearly were not absolutely invariant, they also did not exhibit relative invariance: In relative terms, too, the pedal release occurred earlier at the fast than at the slow tempo. There was also variation in PRT% across positions. PDTs varied with tempo and position, and for these times there was also a tempo by position interaction, with tempo effects being much larger in some bars (3-1, 23-1) than in others (7-1, 15-1). PDT%, on the other hand, did not vary with tempo, although it differed across positions and showed a position by tempo interaction. PCTs did not show a tempo main effect, but they varied across positions and exhibited a tempo by position interaction: It can be seen in Figure 7 that they decreased with tempo in some bars (3-1, 19-1) but increased with tempo in others (1-1R, 7-1R). Similarly, PCT% did not vary with tempo but did vary across positions and showed a position by tempo interaction. In sum, then, these data are more complex than the simple hypotheses of absolute versus relative invariance would predict.

BHR's IOIs varied with tempo and across positions, being somewhat longer in bars 3-1, 3-1R, and 19-1 than in bars 7-1, 7-1R, 11-1, and 15-1, and longest in bar 23-1. His preceding tone key releases and pedal releases occurred much earlier than LPH's, around the onset of the IOI and shortly afterwards, respectively. KOTs did not vary with tempo but did differ across positions. PRTs varied significantly with tempo and across positions. PDTs were rather variable; they either immediately preceded the onset of the second tone or followed it, especially in bar 7-1(R). Therefore, the means shown in Figure 7 are not always representative, and the PDTs were not subjected to statistical analysis. However, it is clear from Figure 7 that they varied with tempo and that they were not relationally invariant, at least not in bars 7-1(R). It is equally evident that PDTs were not absolutely invariant. However, PRT% did not vary significantly with tempo, and PDT%, when the pedal depressions were very close to second tone onsets, did not either. Therefore, PCT% must also have been approximately invariant in those cases. BHR's data then lend some support to the relational invariance hypothesis, in as much as pedal events seemed to be partially coordinated with tone onsets.

Other analysis sets. The remaining four analysis sets, whose results are presented in the Appendix, were bars 3-5, 3-8, 4-1, and 4-3, and their corresponding positions in the music. The data were less complete in these sets, due to the paucity or absence of pedal events in some positions. In the final analysis set, some completely uninformative positions were replaced with adjacent positions.

Summary of statistical analyses. Tables 2-4 summarize the significance levels of the F-tests in the 2-way ANOVAs. No adjustment was made for the number of tests conducted, since the pattern of the data seemed more important than the significance levels of particular effects. Table 2 shows the main effects of (global) tempo, Table 3 those of position (i.e., local tempo), and Table 4 the interactions between these two factors. "No test" indicates missing or insufficient data.

Table 2. *Summary of tempo main effects (i.e., global tempo effects).*

Anal. set	IOI	Pedal release time		Pedal depression time		Pedal change time	
LPH		PRT	PRT%	PDT	PDT%	PCT	PCT%
2-2 etc.	**	**	*	**			*
3-1 etc.	**	**	****	**			
3-5 etc.	*						
3-8 etc.	***	*		***	**		
4-1 etc.	**						
4-3 etc.	**	*		*			
BHR							
2-2 etc.	****						
3-1 etc.	***	*		no test	no test	no test	no test
3-5 etc.	****			*		no test	no test
3-8 etc.	****	no test	no test	no test	no test	no test	no test
4-1 etc.	**	*		**		**	
4-3 etc.	****						

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

Table 3. *Summary of position main effects (i.e., local tempo effects).*

Anal. set	IOI	Pedal release time		Pedal depression time		Pedal change time	
LPH		PRT	PRT%	PDT	PDT%	PCT	PCT%
2-2 etc.	***	****	***	***		**	****
3-1 etc.	****	****	****	****	****	****	****
3-5 etc.	****	**	*	***		*	
3-8 etc.	****			**	***		
4-1 etc.	****	**	****	****	***	****	****
4-3 etc.	****	****	****	****	****	****	****
BHR							
2-2 etc.	****		**	**	***		
3-1 etc.	****	****		no test	no test	no test	no test
3-5 etc.	***	****	****	****	****	no test	no test
3-8 etc.	****	no test	no test	no test	no test	no test	no test
4-1 etc.	****	****	****				
4-3 etc.	****	****	****	****	****		

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

Table 4. *Summary of tempo by position interactions.*

Anal. set	IOI	Pedal release time		Pedal depression time		Pedal change time	
LPH		PRT	PRT%	PDT	PDT%	PCT	PCT%
2-2 etc.							
3-1 etc.					**	**	*
3-5 etc.							
3-8 etc.							
4-1 etc.				**	*	**	**
4-3 etc.				*			
BHR							
2-2 etc.	*		*				
3-1 etc.				no test	no test	no test	no test
3-5 etc.						no test	no test
3-8 etc.		no test	no test	no test	no test	no test	no test
4-1 etc.		*	*				
4-3 etc.							

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

The IOI columns in the tables indicate that IOI duration always varied with global tempo (Table 2), as it should, and also across positions (Table 3), which demonstrates that local tempo variation was present in the data, even across positions that were structurally as similar as possible. Table 4 shows that tempo by position interactions were generally absent, which indicates that IOI duration was very nearly relationally invariant in these performances (cf. Repp, 1994a).

For the pedaling times, Table 2 shows 10 instances in which global tempo affected absolute but not relative times, as predicted by the relative invariance hypothesis, but only one instance where the opposite was the case, as predicted by the absolute invariance hypothesis. In three instances, both measures were affected by global tempo, and in the many remaining instances neither showed an effect. The results for PCT are particularly inconclusive. The absolute invariance hypothesis thus can be rejected for both PRT and PDT; relational invariance seems to hold occasionally for these events, but not always.

Table 3 shows that both absolute and relative pedaling times were strongly affected by local tempo in many instances. Thus, both the absolute and relative invariance hypotheses must be rejected with regard to local tempo. The high significance levels of many effects indicate that pedal timing exhibited reliable changes as a

function of local tempo variation, but their pattern was more complex than predicted by any simple invariance hypothesis. Table 4 shows that, furthermore, global and local tempo effects sometimes interacted in a reliable fashion, which provides further evidence for the controlled complexity of pedal timing.

GENERAL DISCUSSION

The present results demonstrate that pedal timing in expressive piano performance does not follow any simple pattern. Yet, the timing pattern is reproducible by the same pianist. Although there is variation in pedal use from one performance to the next, as already noted by Heinlein (1929b) and Banowetz (1985), whenever the pedal *is* used in the same position in the music, its action tends to be timed similarly. Although no direct measures of timing variability across performances of the same nominal tempo were calculated here, the pianists' relative consistency is reflected in the high significance levels of many of the statistical effects (Tables 2-4), all of which relied on a comparison (F ratio) of some main effect or interaction with its variability across performances within the same global tempo category. A more direct impression of the pianists' consistency may be gained by visually comparing the data in the first, third, and seventh compartments, as well as those in the second and fourth compartments, of Figures 6 and 7. They represent replications of the identical musical material *within* performances, and they generally show a very similar pedal timing pattern. (Note that some of the temporal variability in the raw data has been eliminated in these graphs by averaging over the three performances at each global tempo.)

Since nothing was known about precise pedal timing before this study began, the author may be forgiven for entertaining some perhaps simple-minded but heuristically useful hypotheses, for example that pedal release and depression times might be absolutely invariant. These hypotheses can now be safely rejected. Clearly, neither pedal releases nor pedal depressions occur at a fixed time after each key depression. This was especially clear in their variation across positions within (and between!) analysis sets. Only BHR's pedal releases showed some degree of invariance in that they usually occurred within 100 ms after a key depression, but their timing, too, varied significantly across positions. Absolute invariance did not hold across variations in global tempo either, although in some analysis sets there was not sufficient statistical evidence to reject this hypothesis. Even BHR's very early pedal releases, which could not be expected to vary much with global tempo, showed significant tempo effects in two analysis sets.

Perhaps the most promising hypothesis was that of absolute invariance of pedal change times. The rapid sequence of pedal release and depression is a highly overlearned and automatic action pattern, and one might think that it would be executed as quickly as possible without much regard to context or tempo. This proved to be wrong also. Both pianists showed a wide range of pedal change times, which varied especially across positions (compare also Figs. 6 and 7). The behavior of pedal change times in the face of global tempo changes was less clear.

If absolute invariance does not hold, then the hypothesis of relational invariance seems the next most plausible candidate. However, it too finds little support in the present data. The variation across positions (local tempo) provides the most striking counterevidence: All relative pedal timing measures exhibited large position effects, except for BHR's relative pedal change times. It is clear, however, from comparisons across analysis sets that these latter times were not constant either. Only with regard to variations in global tempo does the relational invariance hypothesis find some support: In a number of instances, relative pedal release and depression times did not vary with tempo, though their absolute timing did. However, there are also several counterexamples.

The present study did not attempt to explain the patterns of pedal timing with reference to the structural features of the music that govern expressive manual timing. However, some attention was paid to the timing of the key release for the preceding melody note which, in LPH's case at least, seemed to depend in part on the harmonic relationship of the overlapping tones. Delayed key releases seemed to go with delayed pedal releases and depressions, at least across analysis sets. Within analysis sets, however, there seemed to be no precise coordination of key and pedal releases, even though they tended to occur at roughly the same time in LPH's playing (cf. Figure 7).

A tendency for pedaling events to align themselves with manual events was observed in several instances. In the first analysis set (bar 2-2 etc.), pedal actions nearly coincided with grace note onsets for LPH, and they alternated with grace note onsets for BHR, as if these actions were "in phase" in one case and "out of phase" in the

other. This is reminiscent of the preferred phase relationships found in studies of bimanual coordination in simple repetitive movements (e.g., Haken, Kelso, & Bunz, 1985), but the present situation is different in many ways: The moving body parts are of different sizes (fingers versus foot), the timing is relatively rapid, and neither movement is oscillatory in character (although a pedal change may be considered a single cycle of a potentially continuous maneuver). In several other instances there was a tendency for pedal depressions to coincide with note onsets, and BHR's early pedal releases may be considered aligned with note onsets as well. However, there were also many instances of non-alignment, and the data are basically inconclusive as to whether there were any preferred phase relationships between manual and pedal actions.

There were many differences in pedal use and timing between the two pianists, LPH and BHR. While LPH had undergone professional training as a concert pianist and therefore must have given considerable conscious attention to her pedaling technique at various times in her career, BHR is an amateur who never has given much thought to his pedaling skills. Yet, the various individual differences cannot immediately be attributed to differences in skill level, as individual differences in pedal timing may well exist between pianists on the same level of expertise. That is another issue worthy of further research. One skill-related difference between LPH and BHR, however, was in their key overlap times (see also Repp, 1994a): LPH generally played in a *legatissimo* style, which almost certainly enhanced the beauty of her performances, and which was not within BHR's capabilities. In fact, BHR sometimes exhibited gaps between notes that could have been played *legato*. This difference in manual playing style was the likely cause of the most obvious difference between the two pianists in pedal timing, viz., in the early (BHR) versus late (LPH) occurrence of pedal releases within an IOI.

In summary, the present investigation initiated the study of a little investigated topic, pedal timing, in the most complex, but ecologically most valid, situation imaginable: artistic music performance. The results present a rather complex picture of pianists' pedaling skill whose clarification will require considerable further research. Experiments employing simpler materials in more artificial situations may help unravel the factors that influence pedal timing, but a glimpse of the richness of real-life data is a good antidote to future oversimplification.

REFERENCES

Banowetz, J. (1985). *The pianist's guide to pedaling*. Bloomington, IN: Indiana University Press.

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance* (pp. 76-186). Stockholm: Royal Swedish Academy of Music (Publication No. 39).

Desain, P., & Honing, H. (1994). Does expressive timing in music performance scale proportionally with tempo? *Psychological Research, 56*, 285-292.

Gentner, D. R. (1987). Timing of skilled motor performance: Tests of the proportional duration model. *Psychological Review, 94*, 255-276.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics, 51*, 347-356.

Heinlein, C. P. (1929a). The functional role of finger touch and damper-pedaling in the appreciation of pianoforte music. *Journal of General Psychology, 2*, 462-469.

Heinlein, C. P. (1929b). A discussion of the nature of pianoforte damper-pedaling together with an experimental study of some individual differences in pedal performance. *Journal of General Psychology, 2*, 489-508.

Heinlein, C. P. (1930). Pianoforte damper-pedaling under ten different experimental conditions. *Journal of General Psychology, 3*, 511-528.

Heuer, H. (1991). Invariant relative timing in motor-program theory. In J. Fagard & P. H. Wolff (Eds.), *The development of timing control and temporal organization in coordinated action* (pp. 37-68). Amsterdam: Elsevier.

Marek, C. (1972). *Lehre des Klavierspiels*. Zürich: Atlantis.

Neuhaus, H. (1973). *The art of piano playing*. New York: Praeger.

Newman, W. S. (1984). *The pianist's problems*. New York: Da Capo Press.

Philipp, G. (1984). *Klavier, Klavierspiel, Improvisation*. Leipzig: VEB Deutscher Verlag für Musik.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustical Society of America, 92*, 2546-2568.

Repp, B. H. (1994a). Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study. *Psychological Research, 56*, 269-284.

Repp, B. H. (1994b). On determining the basic tempo of an expressive music performance. *Psychology of Music, 22*, 157-167.

Repp, B. H. (1995). Acoustics, perception, and production of *legato* articulation on the piano. *Journal of the Acoustical Society of America, 97*, 3862-3874

Seashore, C. E. (1938). *Psychology of music*. New York: McGraw-Hill. (Reprinted by Dover Publications, 1967.)

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartók: Studies in motor programming. *Cognitive Psychology, 13*, 326-376.

Shaffer, L. H., Clarke, E. F., & Todd, N. P. (1985). Metre and rhythm in piano playing. *Cognition, 20*, 61-77.

Taguti, T., Ohgushi, K., & Sueoka, T. (1994). Individual differences in the pedal work of piano performance. In A. Friberg, J. Iwarsson, E. Jansson, & J. Sundberg (Eds.), *SMAC 93: Proceedings of the Stockholm Music Acoustics Conference July 28 - August 1, 1993* (pp. 142-145). Stockholm: Royal Swedish Academy of Music.

Todd, N. P. McA. (1992). The dynamics of dynamics: A model of musical expression. Journal of the Acoustical Society of America, *91*, 3540-3550.

FOOTNOTES

Psychology of Music, in press (without the Appendix).

[1]Several comments are in order: (1) Relative invariance of absolute durations (in ms) is the same as absolute invariance of relative durations (percentages or proportions). (2) While the three intervals considered may be all absolutely invariant or all relationally invariant, certain other combinations are impossible; e.g., absolute invariance of PCTs together with relational invariance of PRTs implies that PDTs cannot exhibit either type of invariance (PDT = PRT + PCT). Of course, most of these conceivable "mixed" scenarios are theoretically implausible. (3) Absolute and relative invariance are the easier to distinguish the later in the IOI the relevant pedalling event occurs, because early in the IOI the absolute effect of a proportional adjustment to IOI duration may be vanishingly small.

[2]Key release times of the preceding tone (the C5 in the soprano voice) were not included in this analysis. In part, they are reported in Repp (1994a: Fig. 6, tone pair 4-5). For BHR, they were near the beginning of the IOI, but LPH often held the key down much longer than the notation would suggest, even beyond the onset of the tone terminating the IOI (especially in bar 6-2(R)). In those instances of "finger pedaling," the key release times are clearly irrelevant to pedal timing within the IOI.

APPENDIX: ADDITIONAL ANALYSES

Bar 3-5 and corresponding IOIs. This place in the music is similar to bar 3-1 etc. In bars 3-5, 19-5, and 23-5, it again represents an upward pitch movement (here, of a major third) in the soprano voice, accompanied by longer sustained notes in the other voices. In bars 7-5, 11-5, and 15-5, however, the upward pitch movement (here, of a perfect fourth) has shifted to the alto voice (Figure 2). LPH again carried out a pedal change in nearly all instances (Figure 3); BHR, on the other hand, did so only occasionally in bar 23-5 (Figure 4). In bars 3-5(R) and 19-5, BHR depressed the pedal after a short break in pedal use; in bars 7-5(R), 11-5, and 15-5, he released it and depressed it again only in the following IOI.

The quantitative results are shown in Figure A1. Consider first LPH. Her IOIs naturally varied with tempo, despite an anomaly in bar 3-5R, and also across positions, mainly due to a greatly elongated IOI in bar 23-5. Her PRTs occurred relatively early in the IOI (earlier than in Figure 7) and did not vary significantly with tempo (again in contrast to Figure 7), though they occurred later in bars 7-5(R) and 23-5 than in the others. Key releases for the preceding melody tone, which in all instances formed the dissonant interval of a minor second, occurred also much earlier than in the preceding analysis set and at about the same time as pedal releases, though the two events did not seem to be precisely coordinated. KOTs did not vary significantly with either tempo or position, though the interaction was significant. LPH's PDTs occurred close to the onset of the second tone, but the effect of tempo was not significant, indicating considerable variability; however, they did vary across positions. PCTs also varied only across positions, not with tempo. None of the percentage measures showed a significant tempo effect, leaving these data ambiguous with regard to the relational invariance hypothesis. PRT%, however, varied across positions, being longer especially in bar 7-5(R), whereas PDT% and PCT% did not vary.

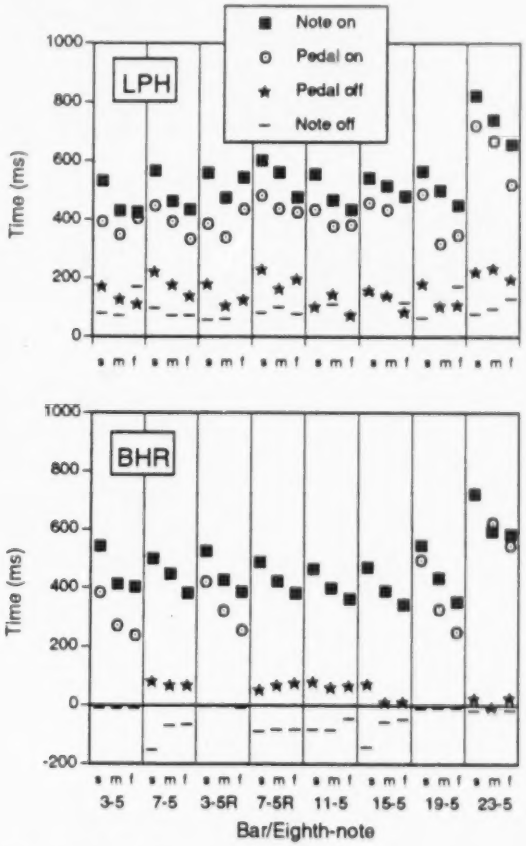


Figure A1. **Pedal timing data for bar 3-5 and corresponding locations.**

BHR's IOIs varied systematically with tempo and also across positions, again mainly due to the slowing down in bar 23-5. Pedal changes occurred only in bar 23-5, where pedal releases and depressions virtually coincided with tone onsets. Pedal releases were found in bars 7-5(R), 11-5, 15-5, and 23-5; they occurred very close to the beginning of the IOI and did not vary with tempo. The main effect of positions was significant, mainly due to the earlier release times in bar 23-5. Preceding note key releases nearly coincided with IOI onset in bars 3-5(R), 19-5, and 23-5. In bars 7-5(R), 11-5, and 15-5, the key for the preceding tone (soprano voice) was released well before the next tone onset, leaving a gap that was inaudible due to pedaling. This was probably caused by BHR's attention to the melodic motive in the alto voice, resulting in a neglect of *legato* in the soprano voice. There was a highly significant effect of position on key overlap, but no effect of tempo. Pedal depressions occurred in bars 3-5(R), 19-5, and 23-5; they did vary with tempo and across positions. Neither PRT% nor PDT% varied with tempo, but they differed strongly across positions, due to earlier releases and later depressions in the later bars, especially bar 23-5. These data are not consistent with any invariance hypothesis, whether absolute or relational.

Bar 3-8 and corresponding IOIs. These positions may be divided into two structurally different sets. In bars 3-8(R), 19-8, and 23-8, melodic upward movement occurs in all four voices, and the pedal helps establish a smooth *legato* transition, which is difficult to achieve through fingering alone. In bars 7-8(R), 11-8, and 15-8, on the other hand, there is downward pitch movement in all active voices; again, the pedal assists with a smooth *legato*. LPH executed pedal changes fairly consistently, though less often in bars 7-8(R) and 11-8 (Figure 3). BHR did so only intermittently; in bars 3-8(R) and 19-8, he had only pedal depressions (Figure 4). The quantitative results are shown in Figure A2.

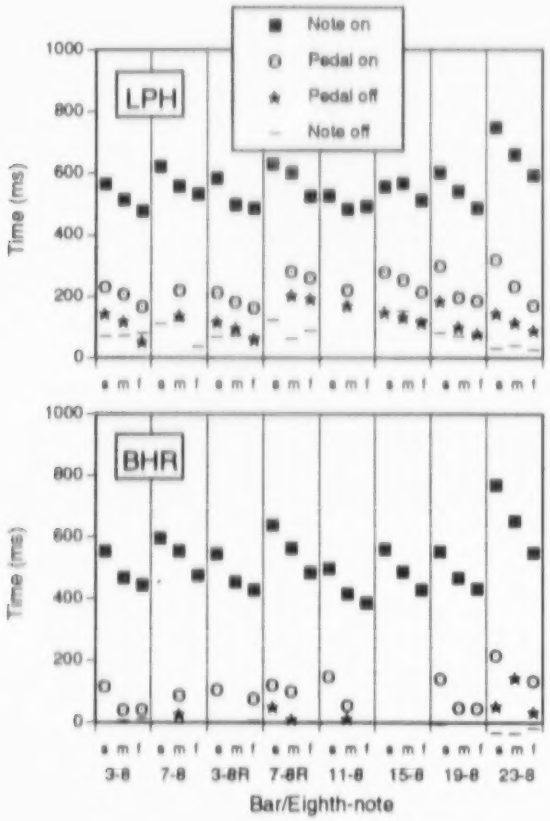


Figure A2. **Pedal timing data for bar 3-8 and corresponding locations.**

It goes without saying that IOI durations varied with tempo and position for both pianists. Preceding note key releases again occurred in the vicinity of pedal releases, though coordination was not precise. They varied significantly across positions, but the effect of tempo was not significant. In LPH's pedaling times, bars 7-8, 7-8R, and 11-8 had to be excluded from analysis due to incomplete data. In the remaining five positions, PRTs varied with tempo, and so did PDTs, but not PCTs. PDTs also varied across positions. Relativization of the timing measures did not eliminate effects of tempo; although the tempo variation of PRT% was not quite significant, PDT% declined very significantly with increasing tempo and also varied across positions. PCT% showed no significant variation. Again, these data do not follow a very clear pattern, although they are consistent with either type of invariance for PCTs. Note, however, that PCTs were much shorter here than in Figure 8. BHR's data were insufficient for a meaningful statistical analysis.

Bar 4-1 and corresponding IOIs. This position is characterized by a melodic upward movement in all instances, similar to bar 3-1 and its analogues. The melodic motion occurs in the soprano voice in bars 4-1(R), 20-1, and 24-1, in the alto voice in bar 8-1(R), and in the tenor voice in bars 8-1 and 16-1, with bar 8-1 being different in that the pitch step is a sixth rather than a third (Figure 2). In contrast to the positions analyzed in the two preceding analysis sets, but in common with the bar 3-1 set, there is no harmonic change here, so that no pedal change was required. LPH nevertheless almost always executed this maneuver (Figure 3), whereas BHR did so only in bars 4-1(R), 20-1, and 24-1; in bars 8-1(R), 12-1, and 16-1, he released the pedal and resumed it only one or two IOIs later (Figure 4).

The quantitative results are shown in Figure A3. For both pianists, IOIs varied significantly with tempo and also with position, even when the greatly elongated bar 24-1 was excluded.

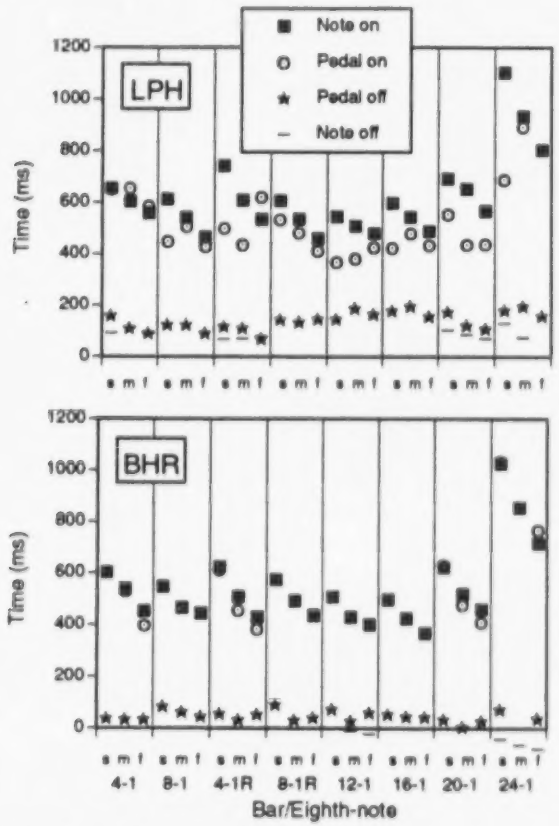


Figure A3. **Pedal timing data for bar 4-1 and corresponding locations.**

For both pianists, the IOIs in bars 4-1(R) and 20-1 were longer than the others, which reflects a subtly different (less ambiguous) rhythmic context in these bars. LPH's pedal releases occurred early in each IOI and did not vary with tempo, though they did vary across positions. Preceding note key releases were evident only in bars 4-1(R), 20-1, and 24-1, where they were again in the vicinity of the pedal releases. They did not vary with either position or tempo. In bars 8-1(R), 12-1, and 16-1, the preceding soprano voice note was released early, so that there was no overlap. This was probably due to fingering (possibly a jump with the fifth finger) and attention to the inner voices. LPH's PDTs were close to the end of the IOI and somewhat irregular, though apparently in a reliable manner: Both the position main effect, with bar 24-1 excluded, and the tempo by position interaction were significant, but the tempo main effect was not. LPH's PCTs likewise varied across positions and interacted with tempo but did not vary with tempo overall. Relativization of PRTs increased their position dependency without introducing a tempo effect. Relativization of PDTs eliminated neither the position main effect nor the tempo by position interaction. It can be seen in Figure 10 that LPH's pedal depressions tended to occur earlier at the fast tempo, but not in all positions. Predictably, PCT% behaved like PDT%, varying both with position and interacting with tempo. In summary, LPH's pedaling times suggest a global tempo independence (absolute invariance) of PRTs, but little more.

BHR's pedal releases occurred immediately after IOI onset, regardless of whether or not a pedal depression followed soon. Despite this apparent uniformity, PRTs varied significantly with position and tempo, and the interaction was significant as well. Preceding note key releases were in evidence only in bars 12-1 and 24-1, and then they preceded IOI onset; in all other cases there were much longer gaps. Pedal depressions, when present (bars 4-1, 4-1R, and 20-1), nearly coincided with the onset of the following tone and therefore clearly varied with tempo but not with position. The same was true for PCTs. PRT% varied strongly across positions, and while the overall tempo dependency disappeared, the interaction remained. PDT% showed no tempo effect, nor did PCT%, although there was a trend towards smaller percentages at faster tempi. Thus, even though the data *look* like a clear example of relative invariance, the actual pattern is more complicated. In other words, even when pedal events are seemingly linked with tone onsets, they are still subject to tempo and position effects.

Bar 4-3 and "corresponding" IOIs. This last vertical section through the score was not strictly vertical, due to the structural heterogeneity of these "corresponding" positions and the resulting absence of pedaling events in some of them. Bars 4-3(R) and 20-3 exhibit a conjoint upward motion in the bass voice; LPH executed a pedal change, whereas BHR only released the pedal. A similar pitch step occurs in bar 8-3(R); LPH changed pedal, but BHR did not do anything consistent. Therefore, his data for bar 8-5(R) were substituted, where a comparable pitch motion occurs in the bass voice, and where he rather consistently executed a pedal change. In bars 12-3 and 16-3, a comparable pitch step occurs in the tenor voice; BHR showed a consistent pedal depression here, but LPH did nothing in this IOI, holding the pedal down through it. Therefore, her pedaling data for the following IOI (bars 12-4 and 16-4) were substituted, where there is a conjoint downward pitch motion in the alto voice, and where she consistently executed a pedal change. Finally, LPH was found to be quite inconsistent in her pedaling during the extra-long IOI in position 24-3, whereas BHR always changed pedal there. LPH's data for bar 24-3 were therefore omitted from analysis; BHR's data for bar 24-3 are shown in Figure A4 but were not included in the statistical analysis because of the extra-long IOI duration.

So, what can be said about this somewhat heterogeneous cross-section of data? IOIs naturally varied with tempo and across positions for both pianists. LPH showed remarkable variation in pedaling. Her PRTs varied with tempo and across positions; the interaction was not significant, even though it seems that there was no tempo dependency in the "imported" bars 12-4 and 16-4. Preceding note key releases were relatively late here, probably due to the harmonic consonance of the overlapping tones (intervals of fourths, fifths, and sixths); note that in bar 8-3(R), and possibly in bars 12-3 and 16-3 as well (depending on fingering), the key overlap was between hands. There was a significant effect of position only on overlap times. PDTs varied with tempo and position, and the interaction was also significant. PCTs did not vary with tempo but did vary across positions. PRT% no longer showed a tempo effect, but the position dependency remained. The same was true for PDT% and for PCT%. Thus, here there is some indication of relational tempo invariance of pedal releases and depressions, together with either

absolute or relative invariance of PCTs. However, the variation of pedal actions across positions is far from simple.

BHR's pedal release times varied across positions but not with tempo. This was equally true for the corresponding percentages. Preceding note key releases were evident only in bars 12-3 and 16-3, where there was in fact some key overlap (between hands); in bar 16-3, pedal depression actually preceded the key release, thus prolonging the preceding tone throughout the IOI. BHR's PDTs were extremely different between the "imported" bar 8-5(R) and bars 12-3 and 16-3, but they did not vary with tempo. PCTs in the latter two bars did not vary with tempo, but neither did the corresponding percentages. Finally, although bar 24-3 was not included in the statistical analysis, it should be noted that pedal depressions in this extra-long IOI occurred no later than those in bar 8-5(R), and pedal releases occurred even earlier. Of course, this bar was also structurally quite different (cf. Figure 2).

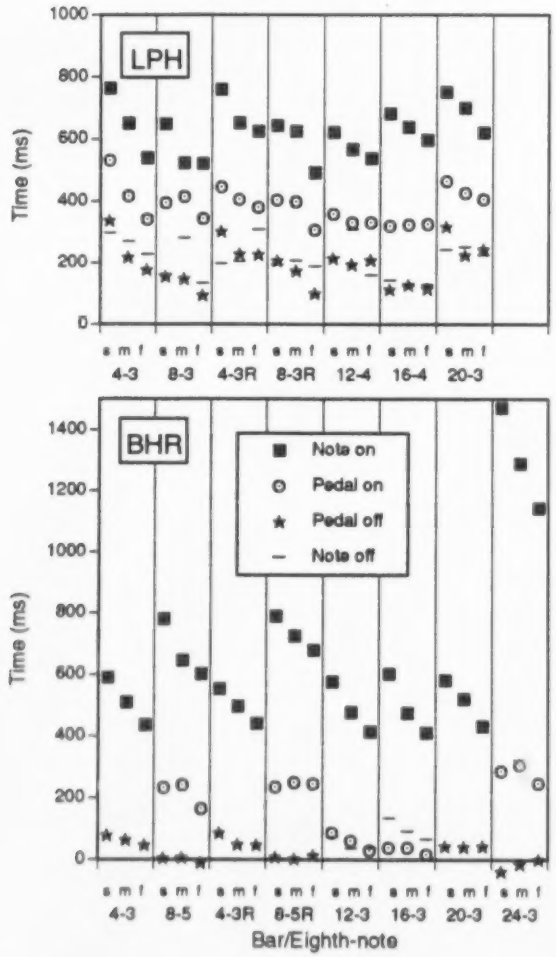


Figure A4. **Pedal timing data for bar 4-3 and corresponding locations, with some substitutions (see text).**

Appendix

SR #	Report Date	NTIS #	ERIC #
SR-21/22	January-June 1970	AD 719382	ED 044-679
SR-23	July-September 1970	AD 723586	ED 052-654
SR-24	October-December 1970	AD 727616	ED 052-653
SR-25/26	January-June 1971	AD 730013	ED 056-560
SR-27	July-September 1971	AD 749339	ED 071-533
SR-28	October-December 1971	AD 742140	ED 061-837
SR-29/30	January-June 1972	AD 750001	ED 071-484
SR-31/32	July-December 1972	AD 757954	ED 077-285
SR-33	January-March 1973	AD 762373	ED 081-263
SR-34	April-June 1973	AD 766178	ED 081-295
SR-35/36	July-December 1973	AD 774799	ED 094-444
SR-37/38	January-June 1974	AD 783548	ED 094-445
SR-39/40	July-December 1974	AD A007342	ED 102-633
SR-41	January-March 1975	AD A013325	ED 109-722
SR-42/43	April-September 1975	AD A018369	ED 117-770
SR-44	October-December 1975	AD A023059	ED 119-273
SR-45/46	January-June 1976	AD A026196	ED 123-678
SR-47	July-September 1976	AD A031789	ED 128-870
SR-48	October-December 1976	AD A036735	ED 135-028
SR-49	January-March 1977	AD A041460	ED 141-864
SR-50	April-June 1977	AD A044820	ED 144-138
SR-51/52	July-December 1977	AD A049215	ED 147-892
SR-53	January-March 1978	AD A055853	ED 155-760
SR-54	April-June 1978	AD A067070	ED 161-096
SR-55/56	July-December 1978	AD A065575	ED 166-757
SR-57	January-March 1979	AD A083179	ED 170-823
SR-58	April-June 1979	AD A077663	ED 178-967
SR-59/60	July-December 1979	AD A082034	ED 181-525
SR-61	January-March 1980	AD A085320	ED 185-636
SR-62	April-June 1980	AD A095062	ED 196-099
SR-63/64	July-December 1980	AD A095860	ED 197-416
SR-65	January-March 1981	AD A099958	ED 201-022
SR-66	April-June 1981	AD A105090	ED 206-038
SR-67/68	July-December 1981	AD A111385	ED 212-010
SR-69	January-March 1982	AD A120819	ED 214-226
SR-70	April-June 1982	AD A119426	ED 219-834
SR-71/72	July-December 1982	AD A124596	ED 225-212
SR-73	January-March 1983	AD A129713	ED 229-816
SR-74/75	April-September 1983	AD A136416	ED 236-753
SR-76	October-December 1983	AD A140176	ED 241-973
SR-77/78	January-June 1984	AD A145585	ED 247-626
SR-79/80	July-December 1984	AD A151035	ED 252-907
SR-81	January-March 1985	AD A156294	ED 257-159
SR-82/83	April-September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022
SR-86/87	April-September 1986	AD A176816	ED 278-066
SR-88	October-December 1986	PB 88-244256	ED 282-278

SR-89/90	January-June 1987	PB 88-244314	ED 285-228
SR-91	July-September 1987	AD A192081	**
SR-92	October-December 1987	PB 88-246798	**
SR-93/94	January-June 1988	PB 89-108765	**
SR-95/96	July-December 1988	PB 89-155329	**
SR-97/98	January-June 1989	PB 90-121161	ED 32-1317
SR-99/100	July-December 1989	PB 90-226143	ED 32-1318
SR-101/102	January-June 1990	PB 91-138479	ED 325-897
SR-103/104	July-December 1990	PB 91-172924	ED 331-100
SR-105/106	January-June 1991	PB 92-105204	ED 340-053
SR-107/108	July-December 1991	PB 92-160522	ED 344-259
SR-109/110	January-June 1992	PB 93-142099	ED 352-594
SR-111/112	July-December 1992	PB 93-216018	ED 359-575
SR-113	January-March 1993	PB 94-147220	ED 366-020
SR-114	April-June 1993	PB 94-196136	ED 371-423
SR-115/116	July-December 1993	PB 95-154936	ED 378-624
SR-117/118	January-June 1994		

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corporation (CMC)
3900 Wheeler Avenue
Alexandria, VA 22304-5110

In addition, Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, CA 92122

**Accession number not yet assigned

Contents—Continued

- Detectability of Duration and Intensity Increments in Melody Tones: A Partial Connection between Music Perception and Performance
  Bruno H. Repp 173
- Acoustics, Perception, and Production of *Legato* Articulation on a Digital Piano
  Bruno H. Repp 193
- Pedal Timing and Tempo in Expressive Piano Performance: A Preliminary Investigation
  Bruno H. Repp 211

Appendix 233

Haskins Laboratories Status Report on

Speech Research

SR-117/118
JANUARY-JUNE 1994

Contents

- The Universality of Intrinsic F0 of Vowels
  D. H. Whalen and Andrea G. Levitt 1
- Intrinsic F0 of Vowels in the Babbling of 6-, 9- and 12-month-old French- and English-learning Infants
  D. H. Whalen, Andrea G. Levitt, Pai-Ling Hsiao, and Iris Smorodinsky 15
- Knowledge from Speech Production Used in Speech Technology: Articulatory Synthesis
  Richard S. McGowan 25
- Nonsegmental Influences on Velum Movement Patterns: Syllables, Sentences, Stress, and Speaking Rate
  Rena A. Krakow 31
- Articulatory Organization of Mandibular, Labial, and Velar Movements During Speech
  H. Betty Kollia, Vincent L. Gracco, and Katherine S. Harris 49
- An Acoustic and Electropalatographic Study of Lexical and Post-lexical Palatalization in American English
  Elizabeth C. Zsiga 67
- The Discriminability of Nearly Merged Sounds
  Alice Faber and Marianna Di Paolo 81
- The Role of Fundamental Frequency in Signaling Linguistic Stress and Affect: Evidence for a Dissociation
  Gerald W. McRoberts, Michael Studdert-Kennedy, and Donald P. Shankweiler 113
- Orthographic Representation and Phonemic Segmentation in Skilled Readers: A Cross-language Comparison
  Ilana Ben-Dror, Ram Frost, and Shlomo Bentin 133
- Expressive Timing in Schumann's "Träumerei": An Analysis of Performances By Graduate Student Pianists
  Bruno H. Repp 141
- Quantitative Effects of Global Tempo on Expressive Timing in Music Performance: Some Perceptual Evidence
  Bruno H. Repp 161

—Continued Inside—

BEST COPY AVAILABLE